AD-A120 660 THE MOMENTS AND DISTRIBUTIONS OF SOME QUANTITIES ARISING FROM RANDOM ARCS ON THE CIRCLE(U) STANFORD UNIV CA DEPT OF STATISTICS F HUFFER 28 SEP 82 TR-326 1/1

UNCLASSIFIED N00014-76-C-0475 F/G 12/1 NL

END

FILMED

DTIC

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

# THE MOMENTS AND DISTRIBUTIONS OF SOME QUANTITIES ARISING FROM RANDOM ARCS ON THE CIRCLE

By

Fred Huffer

TECHNICAL REPORT NO. 326

SEPTEMBER 28, 1982

Prepared Under Contract
N00014-76-C-0475 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director



DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

## ACKNOWLEDGMENT

I wish to thank my adviser, Professor Herbert Solomon, for his guidance and patience. I am also very grateful to him for introducing me to the field of geometrical probability and to the particular topic of this thesis, random arcs on the circle.

I thank Mrs. Carolyn Knutsen for her excellent and always very prompt typing of this manuscript.

## PREFACE

This work concerns random arcs on the circle. Arcs with randomly chosen lengths are placed uniformly on the circumference of a circle and probabilistic questions are asked about the resulting configuration. Emphasis is placed on deriving expressions for the moments and joint moments of quantities such as the number of uncovered gaps, the measure of the uncovered region and the length of the largest gap. Distributions are also obtained for some of these variables.

The subject of random arcs on the circle falls under the general heading of covering problems. Covering problems in two or three dimensions are of greater utility and interest than one dimensional problems like covering the circle. However, multidimensional covering problems are extremely difficult and little progress has been made on them. Much work has been done on one dimensional covering problems.

The results in this thesis are not motivated by any particular application. However, the consideration of random arcs on the circle arises naturally in many contexts. Fisher (1940) noted a connection between certain tests of hypotheses in time series and the covering of the circle by $n$ uniformly placed arcs of length $t$. This connection is rather indirect.

A more direct connection with random arcs is provided by tests of uniformity for directional data. Suppose $n$ points $P_1, P_2, \ldots, P_n$ have been placed on the circumference of a circle and it is desired to test the hypothesis that these points are uniformly distributed. Assume for convenience that the circumference has length one. There are many reasonable test

statistics. Two of these are mentioned below. The choice of test statistic is governed by the alternatives to uniformity which are of greatest interest.

The $n$ points break up the circumference into $n$ disjoint segments. Let $X$ be the arc length of the longest of these segments. It is reasonable to reject the hypothesis of uniformity if $X$ is too large. A more frequently used test is based on the scan statistic. Let $N(x,h)$ be the number of points $P_i$ contained in the arc $(x,x+h]$ where $x+h$ is evaluated mod 1. The scan statistic $N(h)$ is defined as $N(h) = \sup_x N(x,h)$. $N(h)$ is the largest number of points that can be contained in an arc of length $h$. If $N(h)$ is large this may indicate clustering of the points $P_i$ and thus a possible nonuniformity. Information on the scan statistic may be found in Naus (1966) and Cressie (1977).

Now make each point $P_i$ the counterclockwise endpoint of an arc of length $t$. Then $X < t$ if and only if the circumference is completely convered by these $n$ arcs. Let $L(x,t)$ be the number of arcs which cover an arbitrary point $x$ on the circumference. Define the maximal covering $L(t)$ by $L(t) = \sup_x L(x,t)$. Clearly $L(t) = N(t)$. Thus both statistics $X$ and $N(t)$ can be easily interpreted in terms of the random placement of arcs of <u>fixed</u> length on the circumference. See Siegel (1979) for another test of uniformity which can be similarly interpreted.

Problems involving random convex hulls lead naturally to coverage problems involving arcs of <u>random</u> length. Let $Z_1, Z_2, \ldots, Z_n$ be points on the plane generated independently according to the distribution $Q$. Consider the event $B$ that the convex hull of the points $Z_1, \ldots, Z_n$ contains a given convex set $K$. Jewell and Romano (1982) show that each

point $Z_i$ can be mapped into an arc $A_i$ on the unit circle in such a way that $B$ occurs if and only if the arcs $A_1, A_2, \ldots, A_n$ completely cover the circle. If $K$ is a circle centered at the origin and $Q$ has circular symmetry, then it can be shown that the arcs $A_i$ are uniformly placed on the circumference with lengths independent of their positions. This is the standard setting for random arcs on the circle.

Another example where arcs of random length arise is given by Siegel (1978a). The problem of covering a planar region by uniformly placed disks of fixed radius leads to a related circle covering problem with arcs of random length.

There is a fairly substantial literature dealing with random arcs on the circle and related issues such as the covering of line segments or the partitioning of intervals by random points, etc. Most of this is not central to the development which follows. The most relevant history is summarized below. For a more detailed history and bibliography see Siegel (1977) or Solomon (1978).

Stevens (1939) was the first to correctly obtain the probability that $n$ uniformly placed arcs of length $t$ completely cover the circumference of a circle. He also derived the distribution of the number of uncovered gaps on the circumference. Siegel (1978b) derived the moments and distribution of the measure of the uncovered portion of the circumference. This quantity will be denoted by $V$. Votaw (1946) had obtained the distribution of the measure of the covered region in the related problem of covering the line. Siegel (1978a) also gave a general expression for the moments of $V$ when the arcs have random lengths chosen independently from a distribution $F$. The corresponding expression for the distribution

of the number of uncovered gaps was derived by Holst and Siegel (1982). This thesis builds directly on the work of Siegel and Holst.

Some comments are in order on the origins of the ideas found most useful in this paper. The most frequently used notion is that of computing the p-th moment of the measure of a random set $A$ as the probability that $p$ independent uniformly generated points all lie in $A$. Variants of this idea are used for computation of joint moments in situations involving two random sets $A$ and $B$. This idea was formalized and made rigorous by Robbins (1944) who used it to calculate the low order moments of the measure of the region covered by random translates of a fixed set. This method was applied much earlier in other contexts. For elementary applications of the method by M. Crofton see the fifth chapter of Solomon's 1978 monograph.

An argument due to Holst (1980) also proved to be very useful. Expressions for the distributions of $V$ and $V^*$ (defined in section 8) were obtained via this approach. Holst's argument was also used to provide alternative derivations for some of the moments and joint moments considered.

Sections 0 and 1 of this thesis contain definitions and various facts necessary for later calculations.

Section 2 deals with the case of arcs having a fixed nonrandom length. Moments are obtained for the number $G$ of uncovered gaps. The factorial moments of $G$ are found to be especially simple in form. These moments are equivalent to and offer a convenient summary of the distribution of $G$ found by Stevens (1939). Next are presented a number of derivations of the moments of the measure $V$ of the uncovered region.

One derivation is due to Holst (1980). All the arguments are different from the original method of Siegel (1978b). Finally, several derivations are given of the joint moments of $V$ and $G$.

In section 3 the case of random arc lengths is considered. An expression for the factorial moments of $G$ is obtained by the argument of Holst and Siegel (1982). Then Siegel's (1978a) derivation of a general expression for the moments of $V$ is presented. Combining the ideas of the two previous results leads to an expression for the joint moments of $G$ and $V$. These joint moments are evaluated in the case where the $n$ arcs have lengths sampled independently from a distribution of the form $F(x) \propto (x-\alpha)_+^\beta$ for $0 \leq x \leq 1$.

Section 4 treats the special case where the arc lengths have the uniform distribution $F(x) = x$. Many special arguments are applicable in this case. The $2n$ points which are the endpoints of the $n$ random arcs divide the circumference into $2n$ segments. Let $G_k$ be the number of segments covered exactly $k$ times. $V_k$ is the measure of that part of the circumference covered exactly $k$ times. It is shown that the conditional distribution $\mathcal{L}(V_k|G_k=j)$ is Beta$(j,2n-j)$. Formulas for $EV_k$, $EG_k$, $EV_k^2$ and $E\binom{G_k}{2}$ are obtained for all $k$. $E\binom{G_1}{p}$ and $EV_1^p$ are also calculated for all $p$.

Section 5 considers the random variables $G_k$ and $V_k$ when the arc lengths have an arbitrary distribution $F$. General expressions are given for the quantities $EV_k$, $EG_k$, $EV_jV_k$ and $EG_jG_k$.

Section 6 deals with the spacings $S_1,S_2,\ldots,S_n$ between $n$ independent uniformly distributed points on the circumference. A recursion is developed allowing one to compute the joint distribution of $\Sigma_{i=1}^{k+j} S_i$ and $\Sigma_{i=1}^{k} S_i + \Sigma_{i=1}^{m} S_{k+j+i}$.

In section 7 expressions for $P\{G=K\}$ and the conditional moments $E(V^p|G=K)$ are derived as corollaries of the results in section 3.

Section 8 contains the development of the distribution and moments of a quantity $V^*$ which is an upper bound for $V$.

Let $H_k$ denote the length of the k-th largest uncovered gap on the circumference. The distribution of $H_k$ is derived in section 9. It is an immediate corollary of the distribution of $G$ obtained by Holst and Siegel (1982) and presented in section 7. An upper bound is also given for the moments of $H_1$.

In section 10 a general expression for the cumulative distribution of $V$ is derived using the argument of Holst (1980). The distribution of $V$ is found explicitly when $F(x) \propto (x-a)_+^\beta$.

TABLE OF CONTENTS

# THE MOMENTS AND DISTRIBUTIONS OF SOME QUANTITIES ARISING FROM RANDOM ARCS ON THE CIRCLE

By
Fred Huffer

## 0. Description and Notation.

Consider the problem of tossing $n$ arcs at random on the circumference of a circle. More precisely, choose $n$ points $P_1, P_2, \ldots, P_n$ uniformly and independently from a circle of unit circumference. Take the arc lengths $L_1, L_2, \ldots, L_n$ to be i.i.d. from the distribution $F$. Then let $P_i$ serve as a counterclockwise endpoint of an arc with length $L_i$.

Let $P_{(1)}, P_{(2)}, \ldots, P_{(n)}$ be the clockwise ordering of the points with $P_1 = P_{(1)}$. Defnine $S_k = P_{(k+1)} - P_{(k)}$ where $P_{(n+1)} = P_{(1)}$. $S_k$ is the clockwise distance from $P_{(k)}$ to $P_{(k+1)}$. $S_1, S_2, \ldots, S_n$ are the spacings between the counterclockwise endpoints of the arcs. (Warning: The spacings between points chosen uniformly and independently on the circumference will always be denoted by $S_i$ even if they are not the spacings generated by the points $P_1, P_2, \ldots, P_n$.)

The vacancy is that part of the circumference which is not covered by any arc. Let $V$ denote the length (or measure) of the vacancy. $G$ will denote the number of uncovered gaps on the circumference. The vacancy consists of $G$ distinct segments. This report will be concerned with the exact calculation of the moments of $V$ and $G$.

## 1. Preliminary Facts.

Before proceeding it is necessary to list certain facts about the distribution of the spacings $S_1, S_2, \ldots, S_n$ produced by tossing $n$ points at random on the circumference of a circle. These facts are well known and are included only for the sake of completeness.

(1.1) Fundamental lemma:

$$P\{S_1 > a_1,\ldots,S_n > a_n\} = P\{S_1 > \sum_{i=1}^{n} a_i\} = (1 - \sum_{i=1}^{n} a_i)_+^{n-1} \quad .$$

Proof. Let $X_1, X_2, \ldots, X_{n-1}$ be i.i.d. uniform on $[0,1]$. Denote the order statistics by $X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}$. Define $S_k = X_{(k)} - X_{(k-1)}$ for $1 \le k \le n$ where $X_{(0)} \equiv 0$ and $X_{(n)} \equiv 1$. It is clear that $(S_1, S_2, \ldots, S_n)$ so defined has the same joint distribution as the spacings considered above. The joint probability density of $(X_{(1)}, \ldots, X_{(n-1)})$ is given by

$$(n-1)!\ I_{\{0 \le X_{(1)} \le X_{(2)} \le \cdots \le X_{(n-1)} \le 1\}} \cdot$$

Let

$$A = \{(X_{(1)}, \ldots, X_{(n-1)}) : S_1 > a_1, \ldots, S_n > a_n\}$$

and

$$B = \{(X_{(1)}, \ldots, X_{(n-1)}) : S_1 > \sum_{i=1}^{n} a_i\} \ .$$

Define the transformation $T$ so that $T(u_1, \ldots, u_{n-1}) = (v_1, \ldots, v_{n-1})$ with $v_i = u_i + \sum_{j=i+1}^{n} a_j$ for $1 \le i \le n-1$. $T$ is a 1-1 mapping with Jacobian equal to 1. $T$ maps $A$ onto $B$ so that $P(A) = P(B)$. It is easy to see that $P(B) = (1 - \sum_{i=1}^{n} a_i)_+^{n-1}$ .

The following useful facts are immediate corollaries of the previous lemma.

(1.2) The joint distribution of $(S_1, \ldots, S_n)$ is exchangeable:

$$\mathcal{L}(S_1, \ldots, S_n) = \mathcal{L}(S_{\tau(1)}, \ldots, S_{\tau(n)}) \text{ for all permutations } \tau \ .$$

$$(1.3)\qquad P\{S_1 > a,\ S_2 > a,\dots,S_k > a\} = (1-ka)_+^{n-1} \quad \text{for} \quad k \le n .$$

$$(1.4)\qquad \mathcal{L}((S_1-a_1)_+,\dots,(S_n-a_n)_+ \mid S_1 > a_1,\dots,S_n > a_n)$$

$$= \mathcal{L}((1-\sum_{i=1}^{n} a_i)(S_1,\dots,S_n)) \quad \text{with the special case}$$

$$\mathcal{L}((S_1-a)_+,\dots,(S_k-a)_+ \mid S_1 > a,\dots,S_k > a)$$

$$= \mathcal{L}((1-ka)(S_1,\dots,S_k)) \quad \text{for all} \quad 1 \le k \le n .$$

It is also necessary to know the Dirichlet integral.

$$(1.5)\qquad E(\prod_{j=1}^{n} S_j^{\alpha_j-1}) = \left[\frac{\Gamma(n)\prod_{j=1}^{n}\Gamma(\alpha_j)}{\Gamma(\sum_{j=1}^{n}\alpha_j)}\right]$$

if $\alpha_j > 0$ for $1 \le j \le n$ .

## 2. The Case of Constant Arc Length.

Using the results of the previous section it is straightforward to calculate the moments of $G$ and $V$ when $F(x) = I_{\{x \ge a\}}$ so that the arc length is constant $= a$.

The number of uncovered gaps is

$$(2.1)\qquad G = \sum_{i=1}^{n} I_{\{S_i > a\}} .$$

Therefore

$$EG^p = \sum_{i_1,\dots,i_p} P\{S_{i_1} > a,\dots,S_{i_p} > a\} .$$

Grouping the terms according to the number of distinct indices $i_k$ and using exchangeability (1.2) this becomes

$$= \sum_{k=1}^{p\wedge n} S(p,k) \frac{n!}{(n-k)!} P\{S_1 > a,\ldots,S_k > a\} .$$

Now use (1.3)

$$(2.2) \qquad = \sum_{k=1}^{p\wedge n} S(p,k) \frac{n!}{(n-k)!} (1-ka)_+^{n-1} = EG^p .$$

Here $S(p,k)$ equals the number of ways to partition a set containing $p$ elements into $k$ nonempty subsets. These are called Stirling numbers of the second kind.

A more aesthetic formula is obtained by considering $\binom{G}{p}$ instead of $G^p$. Let $\sigma = \{\sigma_1,\sigma_2,\ldots,\sigma_p\}$ denote a subset of $\{1,2,\ldots,n\}$. Define $T = \{\sigma: |\sigma| = p\}$ where $|\sigma|$ is the cardinality of $\sigma$. Clearly

$$\binom{G}{p} = \sum_{\sigma\in T} I\{S_{\sigma_1} > a,\ldots,S_{\sigma_p} > a\}$$

so that by exchangeability (1.2) and (1.3) it follows that

$$(2.3) \qquad E\binom{G}{p} = \binom{n}{p} P\{S_1 > a,\ldots,S_p > a\} = \binom{n}{p}(1-pa)_+^{n-1} .$$

The special cases of most interest are

$$EG = n(1-a)_+^{n-1} \quad \text{and}$$

$$EG^2 = n(1-a)_+^{n-1} + n(n-1)(1-2a)_+^{n-1} .$$

---

The measure of the vacancy is given by

(2.4) $$V = \sum_{i=1}^{n} (S_i - a)_+ \ .$$

Using the multinomial expansion gives

(*) $$EV^p = \sum_{\sigma} \binom{p}{\sigma_1, \ldots, \sigma_n} E\left(\prod_{j=1}^{n} (S_j - a)_+^{\sigma_j}\right)$$

where $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ and the sum ranges over $\{\sigma : \Sigma\sigma_i = p\}$ . Note that $\sigma$ is being used in an entirely different way than in the previous argument.

Let $B(\sigma) = \{j : \sigma_j > 0\}$, $b(\sigma) = |B(\sigma)|$, and $A(\sigma) = \bigcap_{j \in B(\sigma)} \{S_j > a\}$. Then

$$E\left(\sum_{j=1}^{n} (S_j - a)_+^{\sigma_j}\right) = P(A(\sigma))\ E\left(\prod_{j=1}^{n} (S_j - a)_+^{\sigma_j} \middle| A(\sigma)\right)$$

since $\prod_{j=1}^{n} (S_j - a)_+^{\sigma_j} = 0$ outside the set $A(\sigma)$ .

$$= (1 - ab(\sigma))_+^{n-1} \{(1 - ab(\sigma))^p\ E(\prod_{j=1}^{n} S_j^{\sigma_j})\}$$

by (1.3) and (1.4) .

$$= (1 - ab(\sigma))_+^{n+p-1} \frac{(n-1)!}{(p+n-1)!} \prod_{j=1}^{n} \sigma_j!$$

using the Dirichlet integral (1.5) .

Substituting this result in (*) and simplifying gives

$$EV^p = \binom{p+n-1}{p}^{-1} \sum_{\sigma} (1-ab(\sigma))_+^{n+p-1} .$$

By easy combinatorics $|\{\sigma : b(\sigma) = k\}| = \binom{n}{k}\binom{p-1}{k-1}$ for $k \leq n \wedge p$. (How many ways can $p$ indistinguishable balls be placed in $n$ boxes so that exactly $k$ boxes are nonempty? First choose the $k$ nonempty boxes in $\binom{n}{k}$ ways. Then distribute the $p$ balls into these $k$ boxes in $\binom{p-1}{k-1}$ ways.) Therefore

$$EV^p = \binom{p+n-1}{p}^{-1} \sum_{k=1}^{n\wedge p} \binom{n}{k}\binom{p-1}{k-1}(1-ka)_+^{n+p-1} . \tag{2.5}$$

This formula was first obtained by Siegel. For the cases $p = 1$ and $p = 2$ the formula becomes

$$EV = (1-a)_+^n \quad \text{and} \quad EV^2 = \frac{2}{n+1}(1-a)_+^{n+1} + \frac{n-1}{n+1}(1-2a)_+^{n+1} .$$

---

A second derivation (due to Holst) of the moments of $V$ is given below. This method of proof can also be used to calculate the distribution of $V$.

Define $I_j = I_{\{S_j > a\}}$. Clearly

$$1 = \sum_{\sigma} \left(\prod_{j\in\sigma} I_j\right)\left(\prod_{k\notin\sigma} (1-I_k)\right)$$

where $\sigma$ ranges over all $2^n$ subsets of $\{1,2,\ldots,n\}$. Multiply both sides of this equation by $V^p$ and take expectations. Grouping the terms according to the cardinality of $\sigma$ and using the exchangeability (1.2) gives

$$EV^p = \sum_{j=0}^{n} \binom{n}{j} E(V^p I_1 \cdots I_j (1-I_{j+1}) \cdots (1-I_n)) .$$

Now substitute for $V$, expand the product of indicator functions and use (1.2) to obtain

$$= \sum_{j=1}^{n} \binom{n}{j} \sum_{k=0}^{n-j} (-1)^k \binom{n-j}{k} E(I_1 \cdots I_{j+k} [\sum_{i=1}^{j} (S_i - a)_+]^p)$$

$$= \sum_{j=1}^{n} \binom{n}{j} \sum_{k=0}^{n-j} (-1)^k \binom{n-j}{k} P\{I_1 \cdots I_{j+k} = 1\}$$

$$\cdot E([\sum_{i=1}^{j} (S_i - a)_+]^p | I_1 \cdots I_{j+k} = 1) .$$

Now apply (1.3) and (1.4) ,

$$= \sum_{j=1}^{n} \binom{n}{j} \sum_{k=0}^{n-j} (-1)^k \binom{n-j}{k} (1-(j+k)a)_+^{n-1}$$

$$\cdot \{(1-(j+k)a)_+^p E(S_1 + \cdots + S_j)^p\} .$$

$(S_1, \ldots, S_n)$ has a Dirichlet distribution with all parameters equal to one. Thus by the "clumping" property $S_1 + \cdots + S_j \sim \text{Beta}(j, n-j)$ so that

$$E(S_1 + \cdots + S_j)^p = \binom{j+p-1}{p} / \binom{n+p-1}{p} .$$

This result can also be obtained by using the multinomial expansion followed by the Dirichlet integral (1.5). Therefore

$$EV^p = \binom{n+p-1}{p}^{-1} \sum_{j=1}^{n} \binom{n}{j} \sum_{k=0}^{n-j} (-1)^k \binom{n-j}{k}\binom{j+p-1}{p}$$

$$\cdot\ (1-(j+k)a)_+^{n+p-1} .$$

$$= \binom{n+p-1}{p}^{-1} \sum_{\ell=0}^{n} (1-\ell a)_+^{n+p-1} \binom{n}{\ell} \sum_{j=1}^{\ell} (-1)^{\ell-j} \binom{\ell}{j}\binom{j+p-1}{p}$$

$$= \binom{n+p-1}{p}^{-1} \sum_{\ell=1}^{n\wedge p} (1-\ell a)_+^{n+p-1} \binom{n}{\ell}\binom{p-1}{\ell-1}$$

since by generating function arguments

$$\sum_{j=1}^{\ell} (-1)^{\ell-j} \binom{\ell}{j}\binom{j+p-1}{p} = \binom{p-1}{\ell-1} \quad \text{for} \quad 1 \le \ell \le p \quad \text{and}$$

zero otherwise (match coefficients for $x^{\ell-1}$ in the identity $(1+x)^{\ell}(1+x)^{-(p+1)} = (1+x)^{\ell-p-1}$) .

---

A third argument for (2.5) will now be given. This derivation introduces ideas which will be used in later sections.

Let $W$ denote the vacancy so that $V$ is the measure of the random set $W$. Choose points $Q_1,\ldots,Q_m$ distributed independently and uniformly on the circumference and independent of $P_1,\ldots,P_n$. Define $B = \bigcap_{j=1}^{m} \{Q_j \in W\}$. It follows by independence that

(2.6) $$P(B|P_1,\ldots,P_n) = (P\{Q_1 \in W|P_1,\ldots,P_n\})^m = V^m .$$

Therefore

(2.7) $$P(B) = EV^m .$$

Define $\underset{\sim}{P} = \{P_1,\ldots,P_n\}$ and $\underset{\sim}{Q} = \{Q_1,\ldots,Q_m\}$. Choose an element $x$ uniformly at random from $\underset{\sim}{P} \cup \underset{\sim}{Q}$. Let $\underset{\sim}{R} = (R_{(1)},\ldots,R_{(m+n)})$ be the clockwise ordering of the points in $\underset{\sim}{P} \cup \underset{\sim}{Q}$ with $R_{(1)} = x$. Let $\underset{\sim}{C} = (C_1,\ldots,C_{m+n})$ with $C_i = I_{\{R_{(i)} \in \underset{\sim}{P}\}}$. $\underset{\sim}{P}$ and $\underset{\sim}{Q}$ are determined if $\underset{\sim}{R}$ and $\underset{\sim}{C}$ are known. It is intuitively clear that $\underset{\sim}{R}$ and $\underset{\sim}{C}$ are independent and that $\underset{\sim}{C}$ is uniformly distributed over the set $\{\underset{\sim}{C}: \Sigma\, C_i = n\}$ which contains $\binom{m+n}{m}$ elements. This is analogous to the basic result in nonparametric statistics which says that under the appropriate null hypothesis the vector of order statistics and the vector of ranks are independent with the rank vectors having a uniform distribution.

In the argument that follows $m+n+1$ will be equivalent to $1$ when it is used as the value of a subscript. Departing from the notation in section (0) let $S_j = R_{(j+1)} - R_{(j)}$ for $1 \le j \le m+n$. The $S_i$ are the spacings between points in $\underset{\sim}{R}$ (not the spacings between points of $\underset{\sim}{P}$ as formerly defined). B is the event that none of the points $Q_1,\ldots,Q_m$ are covered by an arc so that B occurs if and only if $S_j > a$ for all values of $j$ such that $C_j = 1$ and $C_{j+1} = 0$. More informally B occurs if every block of consecutive points from $\underset{\sim}{Q}$ is preceded by a gap. Define $h(\underset{\sim}{C}) = \sum_{j=1}^{m+n} I_{\{C_j=1,C_{j+1}=0\}}$. Then by the independence of $\underset{\sim}{R}$ and $\underset{\sim}{C}$ and equations (1.2) and (1.3) it follows that $P(B|\underset{\sim}{C} = c) = (1-ah(c))_+^{n+m-1}$. Thus

$$P(B) = \binom{n+m}{m}^{-1} \sum_{c} P(B|\underset{\sim}{C} = c) = \binom{n+m}{m}^{-1} \sum_{k} d_k (1-ka)_+^{m+n-1}$$

where

$$d_k = |\{c : h(c) = k\}| \ .$$

$h(\underset{\sim}{C})$ is the number of separate blocks of ones in $(C_1,\ldots,C_{m+n})$ when the values $C_i$ are arranged in a circle. Using simple combinatorial arguments it can be shown that $d_k = \frac{m+n}{k} \cdot \binom{n-1}{k-1}\binom{m-1}{k-1}$. The factor $\binom{n-1}{k-1}$ arises as

the number of solutions to $n = z_1 + \cdots + z_k$ where the $z_i$ are positive integers. The factor $\binom{m-1}{k-1}$ arises similarly. The factor $\frac{m+n}{k}$ is necessary to properly account for the circular symmetry. Therefore

$$EV^m = \binom{n+m}{m}^{-1} \sum_{k=1}^{m \wedge n} \frac{m+n}{k} \binom{n-1}{k-1} \binom{m-1}{k-1} (1-ka)_+^{n+m-1}$$

which is equivalent to (2.5). This formula is symmetric in $m$ and $n$. Thus $E_n V^m = E_m V^n$ where the subscripts on the expectations specify the number of random arcs. The preceding proof motivates this symmetry since the points $P_1,\ldots,P_n$ and $Q_1,\ldots,Q_m$ play symmetric roles in the proof.

— — —

Define $G^* = \sum_{k=1}^{m+n} I_{\{S_k > a\}}$. Note that $G^*$ is not necessarily equal to $G$ because the $S_k$ are defined differently than in equation (2.1). $G^*$ is distributed as the number of gaps when $n+m$ arcs (of length a) are tossed randomly on the circumference. The previous argument involved conditioning on $\underset{\sim}{C}$. By conditioning on $G^*$ yet another derivation of (2.5) is obtained.

Let $\underset{\sim}{D} = (D_1,\ldots,D_{m+n})$ with $D_k = I_{\{S_k > a\}}$. By exchangeability (1.2) the conditional distribution of $\underset{\sim}{D}$ given $G^* = k$ is uniform on the $\binom{m+n}{k}$ elements of $\{\underset{\sim}{D}: \Sigma D_j = k\}$. Since $\underset{\sim}{R}$ and $\underset{\sim}{C}$ are independent, $\underset{\sim}{D}$ and $\underset{\sim}{C}$ are conditionally independent given $G^*$ and the conditional distribution of $\underset{\sim}{C}$ given $G^* = k$ is uniform on the $\binom{m+n}{m}$ elements of $\{\underset{\sim}{C}: \Sigma C_j = n\}$. Thus the conditional distribution of the pair $(\underset{\sim}{C},\underset{\sim}{D})$ given $G^* = k$ is uniform over the set of $\binom{m+n}{m}\binom{m+n}{k}$ possible values.

The event $B$ occurs if and only if $D_i = 1$ for all values of $i$ such that $C_i = 1$ and $C_{i+1} = 0$. Again this simply means that $B$ occurs if every

block of consecutive points from $\underset{\sim}{Q}$ is preceded by a gap. Thus $P(B|G^* = k) = \nu_k / \binom{m+n}{m}\binom{m+n}{k}$ where $\nu_k$ is the number of pairs $(\underset{\sim}{C},\underset{\sim}{D})$ with $G^* = k$ that lead to the occurrence of B. Combinatorial arguments yield

$$\nu_k = \sum_s \frac{m+n}{s}\binom{m-1}{s-1}\binom{n-1}{s-1}\binom{m+n-s}{k-s} .$$

The summation is over all integers $s$ with the usual convention that $\binom{x}{y} = 0$ if $x < y$ or $y < 0$.

To obtain this formula argue as follows. Remember that $h(\underset{\sim}{C})$ is the number of blocks in the circular arrangement of $(C_1, C_2, \ldots, C_{m+n})$. It was earlier shown that $|\{\underset{\sim}{C}: h(\underset{\sim}{C}) = s\}| = \frac{m+n}{s}\binom{m-1}{s-1}\binom{n-1}{s-1}$. Choose $\underset{\sim}{C}$ with $h(\underset{\sim}{C}) = s$. For B to occur every block of points from $\underset{\sim}{Q}$ must be preceded by a gap but the remaining k-s gaps can be placed arbitrarily in the remaining m+n-s positions. This can be done in $\binom{m+n-s}{k-s}$ ways. Thus

$$|\{(\underset{\sim}{C},\underset{\sim}{D}): h(\underset{\sim}{C}) = s \text{ and } B \text{ occurs}\}| = \frac{m+n}{s}\binom{m-1}{s-1}\binom{n-1}{s-1}\binom{m+n-s}{k-s} .$$

Summing over $s$ gives the result.

Therefore (since $P\{G^* = k\} = P_{n+m}\{G = k\}$)

$$P(B) = EP(B|G^*) = \sum_k \frac{\nu_k}{\binom{m+n}{m}\binom{m+n}{k}} P_{n+m}\{G = k\} .$$

$$\frac{\nu_k}{\binom{m+n}{m}\binom{m+n}{k}} = \binom{m+n}{m}^{-1} \sum_s \frac{m+n}{s}\binom{m-1}{s-1}\binom{n-1}{s-1}\binom{m+n}{s}^{-1}\binom{k}{s} .$$

Multiply this by $P_{n+m}\{G=k\}$, sum over $k$, and interchange the order of summation to obtain

$$P(B) = \binom{m+n}{m}^{-1} \sum_s \frac{m+n}{s} \binom{m-1}{s-1}\binom{n-1}{s-1}\binom{m+n}{s}^{-1} E_{n+m}\binom{G}{s} .$$

Using (2.3) and (2.7) now gives

$$EV^m = \binom{m+n}{m}^{-1} \sum_s \frac{m+n}{s} \binom{m-1}{s-1}\binom{n-1}{s-1}(1-sa)_+^{n+m-1}$$

which is the same version of (2.5) found in the preceding derivation. The subscript $j$ used in $P_j$ and $E_j$ means that the probabilities or expectations are computed assuming the number of random arcs is $j$. When there is no subscript the case of $n$ arcs is assumed.

— — —

Using the method of Holst (the second derivation of (2.5)) it is possible to calculate the joint distribution and joint moments of $G$ and $V$. A convenient form of the joint moments is calculated below. The notation of the second derivation of (2.5) will be used. In the following, indices of summation will range over all the integers with the binomial coefficients $\binom{x}{y}$ taken to be zero if $x < y$ or $y < 0$.

$$E[V^p\binom{G}{q}] = \sum_j \binom{j}{q}\binom{n}{j} E[V^p I_1 \cdots I_j (1-I_{j+1}) \cdots (1-I_n)] .$$

Precisely following the earlier derivation yields:

$$= \binom{n+p-1}{p}^{-1} \sum_k (1-ka)_+^{n+p-1} \binom{n}{k} \sum_j \binom{j}{q}\binom{k}{j}(-1)^{k-j}\binom{j+p-1}{p} .$$

With generating function arguments it can be shown that

$$\sum_j \binom{j}{q}\binom{k}{j}\binom{j+p-1}{p}(-1)^{k-j} = \binom{k}{q}\binom{p+q-1}{k-1} .$$

(Factor out $\binom{k}{q}$ from the sum. Then use the result obtained by matching coefficients of $x^{k-1}$ in the identity $(1+x)^{k-q}(1+x)^{-p-1} = (1+x)^{k-q-p-1}$.)

Therefore

$$(2.8) \qquad E[V^p\binom{G}{q}] = \binom{n+p-1}{p}^{-1} \sum_k (1-ka)_+^{n+p-1}\binom{n}{k}\binom{p+q-1}{k-1}\binom{k}{q}$$

with $k$ ranging from $q$ to $\min(n,p+q)$.

---

A second derivation of (2.8) which uses the ideas and notation of the third proof of (2.5) is now given. (A similar but more general argument is given in Section 3 so that this subsection may be skipped if desired.) The reader should review the definitions of the random variables $\underset{\sim}{P}$, $\underset{\sim}{Q}$, $\underset{\sim}{R}$, $\underset{\sim}{C}$, $\{S_i\}_{i=1}^{m+n}$ and the event $B$ which are given in the earlier derivation.

Equation (2.6) is $V^m = P(B|\underset{\sim}{P})$ so that $E\, V^m f(\underset{\sim}{P}) = E\, I_B f(\underset{\sim}{P})$ for all functions $f(\underset{\sim}{P})$. In particular $E\, V^m\binom{G}{q} = E\, I_B\binom{G}{q}$. Note that $G$ is a function of $\underset{\sim}{P}$ whereas $G^* = \sum_{i=1}^{m+n} I_{\{S_i > a\}}$ is a function of $\underset{\sim}{R}$ and not of $\underset{\sim}{P}$.

For $\sigma \subset \{1,2,\ldots,m+n\}$ define $H_\sigma = \bigcap_{j\in\sigma} \{C_j=1,\ S_j > a\}$. If $B$ is true, then $\{C_j=1,\ S_j > a\}$ occurs if and only if there is an arc

beginning at $R_{(j)}$ which is followed by a gap. Thus when $B$ is true, $H_\sigma$ occurs if and only if there are $|\sigma|$ gaps after the indicated arcs. From this it follows easily that $\sum_{\sigma\in T} I_B I_{H_\sigma} = I_B\binom{G}{q}$ where $T = \{\sigma : |\sigma| = q\}$. Taking expectations yields

$$E\, V^m\binom{G}{q} = \sum_{\sigma\in T} P(B\cap H_\sigma)$$

$$= \binom{n+m}{m}^{-1} \sum_{\sigma}\sum_{c} P(B\cap H_\sigma | \underset{\sim}{C} = c)$$

with $\sigma$ ranging over the elements of $T$. Fix $c = (c_1,\dots,c_{m+n})$ and define $\tau = \{j : c_j = 1\}$ and $\lambda = \{j : j\in\tau$ and $j+1\notin\tau\}$ with the usual convention $m+n+1 = 1$. Note that $|\lambda| = h(c)$ where $h(c)$ is as defined in the third proof of (2.5). Clearly $B\cap H_\sigma \cap \{\underset{\sim}{C} = c\} = \phi$ when $\sigma \not\subset \tau$. If $\sigma\subset\tau$ then given $\{\underset{\sim}{C} = c\}$ the event $B\cap H_\sigma$ occurs if and only if $\varepsilon_j > a$ for all $j\in\sigma\cup\lambda$. Thus by the independence of $\underset{\sim}{R}$ and $\underset{\sim}{C}$ and equations (1.2) and (1.3):

$$P(B\cap H_\sigma | \underset{\sim}{C} = c) = \begin{cases} 0 & \text{if } \sigma\not\subset\tau \\ (1-|\lambda\cup\sigma|a)_+^{n+m-1} & \text{if } \sigma\subset\tau . \end{cases}$$

Therefore

$$E\, V^m\binom{G}{q} = \binom{n+m}{m}^{-1} \sum_{\tau}\sum_{\sigma\subset\tau} (1-|\lambda\cup\sigma|a)_+^{n+m-1}$$

with the summation over $\tau$ and $\sigma$ satisfying $|\tau| = n$, $|\sigma| = q$ and $\sigma\subset\tau$. $\lambda$ is the function of $\tau$ defined by $\lambda = \{j : j\in\tau$ and $j+1\notin\tau\}$.

With $k = |\sigma-\lambda|$ the inner sum may be evaluated to obtain

$$= \binom{n+m}{m}^{-1} \sum_{\tau} \sum_{k} \binom{n-|\lambda|}{k}\binom{|\lambda|}{q-k}(1-(k+|\lambda|)a)_{+}^{n+m-1} .$$

Using a previously given result $|\{\tau: |\lambda| = s\}| = d_s = \frac{m+n}{s}\binom{n-1}{s-1}\binom{m-1}{s-1}$ . Thus the summation becomes

$$\binom{n+m}{m}^{-1} \sum_{s} \sum_{k} \frac{m+n}{s}\binom{n-1}{s-1}\binom{m-1}{s-1}\binom{n-s}{k}\binom{s}{q-k}(1-(s+k)a)_{+}^{n+m-1} .$$

There is no need to specify the ranges of $s$ and $k$ in the sum if the usual conventions for the binomial coefficients are followed. With a bit of algebra this becomes

$$\binom{n+m-1}{m}^{-1} \sum_{s,t} \binom{n}{s}\binom{m-1}{s-1}\binom{n-s}{n-t}\binom{s}{t-q}(1-ta)_{+}^{n+m-1}$$

where $t = s+k$. Juggling the binomial coefficients gives

$$\sum_{s} \binom{n}{s}\binom{m-1}{s-1}\binom{n-s}{n-t}\binom{s}{t-q} = \binom{n}{t}\binom{t}{q} \sum_{s} \binom{m-1}{s-1}\binom{q}{t-s}$$

$$= \binom{n}{t}\binom{t}{q}\binom{m+q-1}{t-1}$$

so that

$$E\, V^{m}\binom{G}{q} = \binom{n+m-1}{m}^{-1} \sum_{t} \binom{n}{t}\binom{t}{q}\binom{m+q-1}{t-1}(1-ta)_{+}^{n+m-1}$$

which agrees with (2.8).

---

The attentive reader will have noticed that the fourth derivation of (2.5) (the one which involved conditioning on $G^*$) was just a camouflaged version of the third derivation. A similar type of camouflage applied to the preceding argument yields a third proof of (2.8). The notation and results of the third and fourth derivations of (2.5) will be used in the following.

In the preceding proof it was shown that $E\, V^m\binom{G}{q} = E\, I_B\binom{G}{q}$. Now condition on both $G^*$ and $\underset{\sim}{C}$.

$$E[I_B\binom{G}{q} \mid G^* = k, \underset{\sim}{C} = c] = \binom{m+n}{k}^{-1} \sum_j \binom{j}{q} |\{\underset{\sim}{D}: G = j \text{ and } B \text{ occurs}\}|$$

since $\underset{\sim}{D}$ given $G^* = k$ and $\underset{\sim}{C} = c$ is uniform on $\{\underset{\sim}{D}: \Sigma D_i = k\}$ which has $\binom{m+n}{k}$ elements. Let $s = h(c)$. Clearly $G = j$ implies $j \geq s$. For ease of discussion say that there is a chunk at $i$ if $D_i = 1$. Requiring that $B$ occur determines the location of $s$ chunks. To make $G = j$ distribute $j-s$ chunks in the remaining $n-s$ locations $i$ with $C_i = 1$. This can be done in $\binom{n-s}{j-s}$ ways. Finally, distribute the last $k-j$ chunks in the $m$ locations $i$ with $C_i = 0$. This can be done in $\binom{m}{k-j}$ ways. Thus $|\{\underset{\sim}{D}: G = j \text{ and } B \text{ occurs}\}| = \binom{n-s}{j-s}\binom{m}{k-j}$. Substituting $i = j-s$ gives

$$\sum_j \binom{j}{q}\binom{n-s}{j-s}\binom{m}{k-j} = \sum_i \binom{s+i}{q}\binom{n-s}{i}\binom{m}{k-s-i} \quad (\dagger) \quad .$$

The well-known combinatorial identity

$$\sum_i \binom{d+i}{e}\binom{a}{i}\binom{b}{c-i} = \sum_i \binom{d}{e-i}\binom{a}{i}\binom{a+b-i}{c-i}$$

can be proved using generating functions. Use this identity to show that $(\dagger) = \sum_i \binom{s}{q-i}\binom{n-s}{i}\binom{m+n-s-i}{k-s-i}$. Dividing by $\binom{m+n}{k}$ and using $\binom{m+n-s-i}{k-s-i}/\binom{m+n}{k} = \binom{k}{s+i}/\binom{m+n}{s+i}$ gives

$$(*) \qquad E(I_B\binom{G}{q} \mid G^* = k, \underset{\sim}{C} = c) = \sum_i \binom{s}{q-i}\binom{n-s}{i}\binom{k}{s+i}/\binom{m+n}{s+i}$$

where $s = h(c)$.

$G^*$ is independent of $\underset{\sim}{C}$. $G^*$ is distributed as the number of gaps $G$ when there are $m+n$ random arcs. Thus

$$E(\binom{G^*}{s+i} \mid \underset{\sim}{C} = c) = \binom{m+n}{s+i}(1-(s+i)a)_+^{m+n-1}$$

by (2.3). Unconditioning in $(*)$ therefore yields

$$E(I_B\binom{G}{q} \mid \underset{\sim}{C} = c) = \sum_i \binom{s}{q-i}\binom{n-s}{i}(1-(s+i)a)_+^{m+n-1} .$$

Now use $P\{\underset{\sim}{C}=c\} = \binom{m+n}{m}^{-1} \frac{m+n}{s}\binom{m-1}{s-1}\binom{n-1}{s-1}$ with $s = h(c)$ to obtain

$$E\, I_B\binom{G}{q} = E\, V^m\binom{G}{q} = \binom{m+n}{m}^{-1} \sum_s \sum_i \frac{m+n}{s}\binom{m-1}{s-1}\binom{n-1}{s-1}\binom{s}{q-i}\binom{n-s}{i}(1-(s+i)a)_+^{m+n-1} .$$

Precisely following the preceding derivation from this point on one again obtains (2.8).

---

The following computation gives a fourth argument for (2.8). As in section (0) let $S_1,\ldots,S_n$ be the spacings generated by the points

$P_1,\ldots,P_n$ so that (2.1) and (2.4) holds. For $\sigma \subset \{1,2,\ldots,n\}$ define $H_\sigma = \bigcap_{j\in\sigma} \{S_j > a\}$. Clearly

$$\binom{G}{q} = \sum_{\sigma\in T} I_{H_\sigma} \quad \text{where} \quad T = \{\sigma : |\sigma| = q\} .$$

Therefore

$$E\, V^p \binom{G}{q} = \sum_{\sigma\in T} E\, V^p I_{H_\sigma} = \binom{n}{q} E\, V^p I_{H_\sigma}$$

where $\sigma = \{1,2,\ldots,q\}$. This follows from the exchangeability (1.2).

$$= \binom{n}{q} P(H_\sigma) E(V^p | H_\sigma) = \binom{n}{q} (1-qa)_+^{n-1} E(V^p | H_\sigma)$$

by (1.3).

It follows from (1.4) that

$$(*) \qquad \mathcal{L}((S_1-a)_+,\ldots,(S_q-a)_+,S_{q+1},\ldots,S_n | H_\sigma)$$

$$= \mathcal{L}((1-qa)(S_1,\ldots,S_n)) .$$

Define $(U_1,\ldots,U_n)$ so that

$$\mathcal{L}(U_1,\ldots,U_n) = \mathcal{L}((S_1-a)_+,\ldots,(S_q-a)_+,S_{q+1},\ldots,S_n | H_\sigma) .$$

$$V = \sum_{i=1}^{n} (S_i-a)_+$$

and thus

$$\mathcal{L}(V|H_\sigma) = \mathcal{L}(\sum_{i=1}^{q} U_i + \sum_{i=q+1}^{n} (U_i-a)_+) .$$

Now use (*)

$$= \mathcal{L}(\sum_{i=1}^{q} (1-qa)S_i + \sum_{i=q+1}^{n} ((1-qa)S_i-a)_+)$$

$$= \mathcal{L}((1-qa)(\sum_{i=1}^{q} S_i + \sum_{i=q+1}^{n} (S_i-b)_+)) \quad \text{with } b = \frac{a}{1-qa} .$$

Thus

$$E(V^p|H_\sigma) = (1-qa)^p \; E[\sum_{i=1}^{q} S_i + \sum_{i=q+1}^{n} (S_i-b)_+]^p$$

and

$$E\, V^p \binom{G}{q} = \binom{n}{q}(1-qa)_+^{n+p-1} \; E[\sum_{i=1}^{q} S_i + \sum_{i=q+1}^{n} (S_i-b)_+]^p .$$

Using the multinomial expansion gives

$$E[\sum_{i=1}^{q} S_i + \sum_{i=q+1}^{n} (S_i-b)_+]^p =$$

$$\sum_{\sigma} \binom{p}{\sigma_1,\ldots,\sigma_n} E(\prod_{i=1}^{q} S_i^{\sigma_i} \prod_{i=q+1}^{n} (S_i-b)_+^{\sigma_i})$$

where $\sigma = (\sigma_1,\ldots,\sigma_n)$ and the sum ranges over $\{\sigma:\Sigma\sigma_i=p\}$ . Let $B(\sigma) = \{j:j > q \text{ and } \sigma_j > 0\}$, $\nu(\sigma) = |B(\sigma)|$, and $A(\sigma) = \bigcap_{j\in B(\sigma)} \{S_j > b\}$. Imitating the calculation in the first derivation of (2.5) yields

$$E\left(\prod_{i=1}^{q} S_i^{\sigma_i} \prod_{i=q+1}^{n} (S_i-b)_+^{\sigma_i}\right)$$

$$= P(A(\sigma))E\left(\prod_{i=1}^{q} S_i^{\sigma_i} \prod_{i=q+1}^{n} (S_i-b)_+^{\sigma_i} \,\Big|\, A(\sigma)\right)$$

$$= (1-b\nu(\sigma))_+^{n-1} \{(1-b\nu(\sigma))_+^{p} E(\prod_{i=1}^{n} S_i^{\sigma_i})\}$$

$$= (1-b\nu(\sigma))_+^{n+p-1} \frac{(n-1)!}{(p+n-1)!} \prod_{i=1}^{n} \sigma_i! \; .$$

Therefore

$$E\left[\sum_{i=1}^{q} S_i + \sum_{i=q+1}^{n} (S_i-b)_+\right]^p = \binom{n+p-1}{p}^{-1} \sum_{\sigma} (1-b\nu(\sigma))_+^{n+p-1} \; .$$

Using $b = \frac{a}{1-qa}$ and substituting in the earlier expression for $E\, V^p\binom{G}{q}$ gives

$$E\, V^p\binom{G}{q} = \binom{n}{q}\binom{n+p-1}{p}^{-1} \sum_{\sigma} (1-(q+\nu(\sigma))a)_+^{n+p-1} \; .$$

By easy combinatorics

$$|\{\sigma : q+\nu(\sigma) = k\}| = \binom{n-q}{k-q}\binom{p+q-1}{k-1} \; .$$

(How many ways can $p$ indistinguishable balls be placed in $n$ boxes so that exactly $\nu(\sigma) = k-q$ of the last $n-q$ boxes are nonempty? First choose the $k-q$ nonempty boxes in $\binom{n-q}{k-q}$ says. Place one ball in each of these $k-q$ boxes and distribute the remaining $p+q-k$ balls in the allowed $k$ boxes in $\binom{p+q-1}{k-1}$ ways.) Thus

$$E\,V^p\binom{G}{q} = \binom{n}{q}\binom{n+p-1}{p}^{-1} \sum_k \binom{n-q}{k-q}\binom{p+q-1}{k-1}(1-ka)_+^{n+p-1} .$$

Since $\binom{n}{q}\binom{n-q}{k-q} = \binom{n}{k}\binom{k}{q}$ this is the same as (2.8).

— — —

The special cases $p = 1$ or $q = 1$ of equation (2.8) can be easily derived by other methods.

Allow the arcs to have a variable length $t$ and define

$$G(t) = \sum_{i=1}^{n} I_{\{S_i > t\}} \text{ and } V(t) = \sum_{i=1}^{n} (S_i - t)_+ .$$

Then $-\frac{dV}{dt} = G(t)$ which suggests the following formal manipulation: $-\frac{d}{dt} E(V^{m+1}) = E(-\frac{d}{dt} V^{m+1}) = (m+1)E(V^m G)$. It is straightforward to rigorize this argument (replace the derivative by a difference quotient and use bounded convergence). Thus differentiation of (2.5) leads to

$$E(V^m G) = \frac{1}{m+1}\binom{m+n}{m+1}^{-1} \sum_k \binom{n}{k}\binom{m}{k-1}(n+m)k(1-ka)_+^{n+m-1}$$

which is equivalent to (2.8) with $q = 1$.

Let $\sigma$ denote a subset of $\{1,2,\ldots,n\}$ and define $T = \{\sigma : |\sigma| = q\}$. A little thought suffices to verify that

$$\sum_{\sigma\in T} \left(\prod_{i\in\sigma} I_{\{S_i > a\}}\right)\left(\sum_{j\in\sigma} (S_j - a)_+\right) = \binom{G-1}{q-1} V .$$

The expectation of the left hand side is

$$= \binom{n}{q} E[(\prod_{i=1}^{q} I_{\{S_i > a\}})(\sum_{j=1}^{q} (S_j - a)_+)]$$

$$= \binom{n}{q} P\{S_1 > a, \ldots, S_q > a\} \; qE[(S_1 - a)_+ | S_1 > a, \ldots, S_q > a]$$

$$= \binom{n}{q} (1-qa)_+^{n-1} \; q[\frac{1}{n} (1-qa)_+] = \binom{n-1}{q-1} (1-qa)_+^n \; .$$

The first two lines follow from (1.2). The third line uses (1.3) and (1.4). Therefore

$$E[\binom{G-1}{q-1} V] = \binom{n-1}{q-1} (1-qa)_+^n \; .$$

This equation is different from (2.8) with $p = 1$ but with some algebra the two equations are seen to be consistent.

## 3. Random Arclength.

Now consider the general case where the arc lengths $L_i$ have an arbitrary distribution F. Somewhat different methods are necessary at this level of generality. For instance the formulas (2.1) and (2.4) have no simple analogues.

A formula for the moments of $G$ has been developed by Holst and Siegel. Their argument follows. Let $A_i$ be the subset of the circumference covered by the $i^{th}$ arc (the arc with endpoint $P_i$ and length $L_i$). Define $H_i = \{P_i \notin \bigcup_{j \neq i} A_j$ and $L_i < 1\}$. $H_i$ occurs when the counter-clockwise endpoint $P_i$ of $A_i$ is not covered by any of the arcs. The gaps and the uncovered $P_i$ are in one-to-one correspondence, $G = \sum_{i=1}^{n} I_{H_i}$.

For $\tau \subset \{1,2,\ldots,n\}$ define $H_\tau = \bigcap_{i\in\tau} H_i$. $|\tau|$ denotes the cardinality of $\tau$. It is easy to see that

$$\binom{G}{q} = \sum_{\tau\in T} I_{H_\tau} \quad \text{with} \quad T = \{\tau : |\tau| = q\}.$$

The events $H_\tau$ in the sum are equally probable (since the arcs $A_j$ are i.i.d.) so that taking expectations yields $E\binom{G}{q} = \binom{n}{q} P(H_\sigma)$ where $\sigma = \{1,2,\ldots,q\}$. $H_\sigma$ is the event that $P_1,P_2,\ldots,P_q$ are uncovered.

Let $P_{(1)},\ldots,P_{(q)}$ be the clockwise ordering of the points $P_1,\ldots,P_q$ with $P_{(1)} = P_1$. Let $L_{(1)},\ldots,L_{(q)}$ be the corresponding ordering of $L_1,\ldots,L_q$. Define $P_{(q+1)} \equiv P_{(1)}$. For $1 \le k \le q$ define $S_k$ to be the clockwise distance from $P_{(k)}$ to $P_{(k+1)}$. Observe that this notation differs from that given in section (0).

Let $D_i$ be the event that $A_i \cap \{P_1,\ldots,P_q\} = \phi$ so that $D_i$ occurs when the $i^{th}$ arc does not cover any of the points $P_1,\ldots,P_q$. Then

$$H_\sigma = \left[\bigcap_{i=1}^{q} \{S_i > L_{(i)}\}\right] \cap \left[\bigcap_{i=q+1}^{n} D_i\right].$$

The random variables $P_1,\ldots,P_n$, $L_1,\ldots,L_n$ are mutually independent with the $L_i$ distributed according to F. Thus

$$P(H_\sigma | P_1,\ldots,P_q) = \left\{\prod_{i=1}^{q} F(S_i)\right\}[P(D_n|P_1,\ldots,P_q)]^{n-q}.$$

Let $C_k$ be the event that the arc $A_n$ lies between $P_{(k)}$ and $P_{(k+1)}$. $D_n = \bigcup_{k=1}^{q} C_k$. The $C_k$ are disjoint.

$$P(C_k|P_1,\ldots,P_q) = EP(C_k|L_n,P_1,\ldots,P_q)$$

$$= E((S_k-L_n)_+|S_k) = \int_0^{S_k} (S_k-\ell)_+ \, dF(\ell)$$

$$= \int_0^{S_k} dF(\ell) \int_\ell^{S_k} dx = \int_0^{S_k} F(x)dx \ .$$

Define $g(s) = \int_0^s F(x)dx$ so that $P(C_k|P_1,\ldots,P_q) = g(S_k)$. Then

$$P(D_n|P_1,\ldots,P_q) = \sum_{k=1}^{q} g(S_k) \ .$$

Therefore

$$\binom{n}{q} P(H_\sigma) = \binom{n}{q} EP(H_\sigma|P_1,\ldots,P_q)$$

(3.1)

$$= \binom{n}{q} E\left(\prod_{i=1}^{q} F(S_i)\left\{\sum_{k=1}^{q} g(S_k)\right\}^{n-q}\right) = E\binom{G}{q}$$

where $S_1,\ldots,S_q$ are the spacings between $q$ points chosen uniformly on the circle and $g(s) \equiv \int_0^s F(x)dx$ .

---

A general expression for the moments of $V$ is due to Siegel. His argument uses the idea of the third derivation of (2.5). Choose points $Q_1,\ldots,Q_p$ distributed independently and uniformly on the circumference and independent of $P_1,\ldots,P_n$. Again let $A_i$ be the subset of the circumference covered by the $i^{th}$ arc. Let $B$ be the event that all the points $Q_i$ lie in the vacancy. $B$ occurs if and only if $\{Q_1,\ldots,Q_p\} \cap (\bigcup_{i=1}^{n} A_i) = \phi$. Repeating the argument for (2.7) gives $EV^p = P(B)$.

Let $D_i$ be the event that $A_i \cap \{Q_1,\ldots,Q_p\} = \phi$. $B = \bigcap_{i=1}^{n} D_i$. The events $D_1,\ldots,D_n$ are conditionally independent and equally probable given $Q_1,\ldots,Q_p$. Thus

$$P(B|Q_1,\ldots,Q_p) = \{P(D_i|Q_1,\ldots,Q_p)\}^n = \{\sum_{k=1}^{p} g(S_k)\}^n$$

as in the derivation of (3.1). $S_1,\ldots,S_p$ are the spacings between $Q_1,\ldots,Q_p$. Therefore

$$(3.2) \qquad P(B) = E\, P(B|Q_1,\ldots,Q_p) = E\, V^p = E\{\sum_{k=1}^{p} g(S_k)\}^n$$

where $S_1,\ldots,S_p$ are the spacings between $p$ independent uniformly chosen points on the circumference and $g(s) \equiv \int_0^s F(x)dx$ .

---

The ideas in the derivations of (3.1) and (3.2) can be combined to obtain an equation for the joint moments $E\binom{G}{q}V^p$. Use the notation from the derivation of (3.2). It follows as in the proof of (2.6) that $P(B|P_1,\ldots,P_n,L_1,\ldots,L_n) = V^p$. $G$ is a function of $P_1,\ldots,P_n,L_1,\ldots,L_n$. Thus

$$E\binom{G}{q}V^p = E[\binom{G}{q}E(I_B|P_1,\ldots,P_n,L_1,\ldots,L_n)] = E\binom{G}{q}I_B .$$

Let $H_\tau$ and $T$ be defined as in the derivation of (3.1).

$$\binom{G}{q} = \sum_{\tau \in T} I_{H_\tau}$$

and hence

$$E\binom{G}{q}I_B = \sum_{\tau\in T} P(B\cap H_\tau) = \binom{n}{q}P(B\cap H_\sigma) \quad \text{with } \sigma = \{1,\ldots,q\} .$$

The events $B\cap H_\tau$ in the sum are equally probable because the arcs $A_j$ are i.i.d. and independent of $Q_1,\ldots,Q_p$. $B\cap H_\sigma$ is the event that none of the points $P_1,\ldots,P_q$, $Q_1,\ldots,Q_p$ are covered by any of the arcs.

Let $\underset{\sim}{R} = (R_1,\ldots,R_{p+q})$ be the clockwise ordering of the points $\{P_1,\ldots,P_q,Q_1,\ldots,Q_p\}$ with $R_1 = P_1$. Define $\xi \subset \{1,2,\ldots,p+q\}$ by $i\in\xi$ if and only if $R_i \in \{P_1,\ldots,P_q\}$. $\xi$ is a random set. Let $\xi = \{\xi_1,\ldots,\xi_q\}$ with $\xi_1 < \xi_2 < \cdots < \xi_q$. Let $L_{(k)}$ be the length of the arc beginning at $R_{\xi_k}$. Denote by $S_i$ the clockwise distance from $R_i$ to $R_{i+1}$ where $R_{p+q+1} \equiv R_1$. Let $D_i$ be the event that $A_i \cap \{P_1,\ldots,P_q,Q_1,\ldots,Q_p\} = \phi$. $D_i$ occurs if $A_i$ does not cover any of the points in $\underset{\sim}{R}$.

Using the above notation

$$B\cap H_\sigma = \left(\bigcap_{k=1}^{q} \{S_{\xi_k} > L_{(k)}\}\right) \cap \left(\bigcap_{j=q+1}^{n} D_j\right) .$$

Knowledge of $\underset{\sim}{R}$ and $\xi$ determines $P_1,\ldots,P_q,Q_1,\ldots,Q_p$. Thus conditioning on $\underset{\sim}{R}$ and $\xi$ is equivalent to conditioning on $P_1,\ldots,P_q$, $Q_1,\ldots,Q_p$. The mutual independence of $P_1,\ldots,P_n,L_1,\ldots,L_n,Q_1,\ldots,Q_p$ implies

$$P(B\cap H_\sigma|\underset{\sim}{R},\xi) = \{\prod_{i=1}^{q} F(S_{\xi_i})\}[P(D_n|\underset{\sim}{R},\xi)]^{n-q} .$$

Arguing as in the proof of (3.1) gives

$$P(D_n|\underset{\sim}{R},\xi) = \sum_{j=1}^{p+q} g(S_j)$$

so that

$$P(B \cap H_\sigma|\underset{\sim}{R},\xi) = \{\prod_{k=1}^{q} F(S_{\xi_k})\}[\sum_{j=1}^{p+q} g(S_j)]^{n-q} .$$

Clearly $\underset{\sim}{R}$ and $\xi$ are independent and hence taking expectations yields

$$P(B \cap H_\sigma|\xi) = E(\prod_{k=1}^{q} F(S_{\xi_k})[\sum_{j=1}^{p+q} g(S_j)]^{n-q})$$

where $S_1,\ldots,S_{p+q}$ are the ordered spacings between $p+q$ points chosen uniformly at random on the circumference. The distribution of $(S_1,\ldots,S_{p+q})$ is exchangeable so that $P(B \cap H_\sigma|\xi)$ does not depend on $\xi$ and thus $P(B \cap H_\sigma|\xi) = P(B \cap H_\sigma)$. Taking $\xi = \{1,2,\ldots,q\}$ then gives

$$E\, V^p\binom{G}{q} = \binom{n}{q} E(\prod_{k=1}^{q} F(S_k)\{\sum_{j=1}^{p+q} g(S_j)\}^{n-q}) \tag{3.3}$$

with $S_1,\ldots,S_{p+q}$ as defined above and $g(s) = \int_0^s F(x)dx$ .

---

The expression (3.3) can be explicitly evaluated for the following class of distributions:

$$F(x) = \lambda(x-a)_+^{\beta} \quad \text{where} \quad 0 \le a < 1, \quad \beta > 0 \quad \text{and } \lambda = (1-a)^{-\beta} .$$

Then $g(x) = \frac{\lambda}{\beta+1}(x-a)_+^{\beta+1}$ and (3.3) becomes

$$E\, V^p\binom{G}{q} = \binom{n}{q} E\{(\prod_{k=1}^{q} \lambda(S_k-a)_+^{\beta})(\sum_{j=1}^{p+q} \frac{\lambda}{\beta+1}(S_j-a)_+^{\beta+1})^{n-q}\}$$

(*)

$$= \binom{n}{q} \frac{\lambda^n}{(\beta+1)^{n-q}} \sum_{\sigma} \binom{n-q}{\sigma_1,\ldots,\sigma_{p+q}} E\{\prod_{k=1}^{q} (S_k-a)_+^{\beta} \prod_{j=1}^{p+q} (S_j-a)_+^{\sigma_j(\beta+1)}\}$$

with summation over all $\sigma = (\sigma_1,\sigma_2,\ldots,\sigma_{p+q})$ satisfying $\Sigma\, \sigma_i = n-q$.

Now define

$$B(\sigma) = \{1,2,\ldots,q\} \cup \{j\colon \sigma_j > 0\}, \quad b(\sigma) = |B(\sigma)|$$

and

$$A(\sigma) = \bigcap_{j\in B(\sigma)} \{S_j > a\} .$$

Imitating the calculation in the first derivation of (2.5) yields

$$E(\prod_{k=1}^{q} (S_k-a)_+^{\beta} \prod_{j=1}^{p+q} (S_j-a)_+^{\sigma_j(\beta+1)})$$

$$= P(A(\sigma))E[\prod_{k=1}^{q} (S_k-a)_+^{\beta+\sigma_k(\beta+1)} \prod_{k=q+1}^{p+q} (S_k-a)_+^{\sigma_k(\beta+1)} | A(\sigma)]$$

$$= (1-ab(\sigma))_+^{p+q-1}\{(1-ab(\sigma))_+^{q\beta+(n-q)(\beta+1)} E(\prod_{k=1}^{q} S_k^{\beta+\sigma_k(\beta+1)} \prod_{k=q+1}^{p+q} S_k^{\sigma_k(\beta+1)})\}$$

$$= (1-ab(\sigma))_+^{p+n\beta+n-1} \frac{\Gamma(p+q)}{\Gamma(p+n\beta+n)} \prod_{k=1}^{q} \Gamma((\beta+1)(\sigma_k+1)) \prod_{k=q+1}^{p+q} \Gamma(1+\sigma_k(\beta+1)).$$

Plugging this expression into (*) gives

$$(3.4) \qquad E\,V^p\binom{G}{q} = \binom{n}{q} \frac{\Gamma(p+q)}{(1-a)^{n\beta}(\beta+1)^{n-q}\,\Gamma(p+n\beta+n)}$$

$$\sum_{\sigma} \binom{n-q}{\sigma_1,\ldots,\sigma_{p+q}} (1-ab(\sigma))_+^{p+n\beta+n-1} \prod_{k=1}^{q} \Gamma((\beta+1)(\sigma_k+1)) \prod_{k=q+1}^{p+q} \Gamma(1+\sigma_k(\beta+1))$$

with the summation over all $\sigma = (\sigma_1,\ldots,\sigma_{p+q})$ satisfying $\Sigma\,\sigma_i = n-q$.

The expression (3.4) simplifies in a few special cases. Letting $\beta \downarrow 0$ gives $F(x) = I_{\{x > a\}}$ in the limit. Thus setting $\beta = 0$ in (3.4) should produce the result for the case of constant arc length. Putting $\beta = 0$ and simplifying yields

$$E\,V^p\binom{G}{q} = \binom{n}{q}\binom{n+p-1}{p+q-1}^{-1} \sum_{\sigma} (1-ab(\sigma))_+^{n+p-1}.$$

A simple combinatorial argument shows that

$$|\{\sigma\colon b(\sigma) = k\}| = \binom{p}{k-q}\binom{n-1}{k-1}.$$

(How many ways can $n-q$ indistinguishable balls be placed in $p+q$ boxes so that exactly $k-q$ of the last $p$ boxes are nonempty? First choose the $k-q$ nonempty boxes in $\binom{p}{k-q}$ ways. Place one ball in each of these $k-q$ boxes and distribute the remaining $n-k$ balls in the allowed $k$ boxes in $\binom{n-1}{k-1}$ ways.) Therefore

$$E\,V^p\binom{G}{q} = \binom{n}{q}\binom{n+p-1}{p+q-1}^{-1} \sum_{k} \binom{p}{k-q}\binom{n-1}{k-1}(1-ka)_+^{n+p-1}$$

which is easily seen to be the same as (2.8).

If $p = 0$ then $b(\sigma) = q$ for all $\sigma$. Thus setting $p = 0$ in (3.4) gives

$$E(\binom{G}{q}) = \frac{n!(1-qa)_+^{n\beta+n-1}}{q(1-a)^{n\beta}(\beta+1)^{n-q}\,\Gamma(n\beta+n)} \sum_\sigma \prod_{k=1}^{q} \frac{\Gamma((\beta+1)(\sigma_k+1))}{\Gamma(\sigma_k+1)}$$

where the summation is over $\{(\sigma_1,\ldots,\sigma_q): \Sigma\,\sigma_i = n-q\}$.

Some slight simplification also occurs upon setting $q = 0$ in (3.4).

$$E\,V^p = \frac{(p-1)!}{(1-a)^{n\beta}(\beta+1)^n\,\Gamma(p+n\beta+n)} \sum_\sigma \binom{n}{\sigma_1,\ldots,\sigma_p}(1-ab(\sigma))_+^{p+n\beta+n-1} \prod_{k=1}^{p} \Gamma(1+\sigma_k+\sigma_k\beta)$$

where the summation is over all $\sigma = (\sigma_1,\ldots,\sigma_p)$ satisfying $\Sigma\,\sigma_i = n$ and now $b(\sigma) = |\{j:\, \sigma_j > 0\}|$ .

## 4. Arclengths With the Uniform Distribution.

In the special case where the arc lengths have the uniform distribution $(F(x) = x)$ simplifications occur which permit the calculation of many quantities of interest by entirely combinatorial means. This simplification results from the following observations.

Let $P$ and $L$ be independent random variables with $P$ chosen uniformly on the circumference and $L$ having the uniform distribution $F(x) = x$ for $0 < x < 1$. $P$ and $L$ together determine a random arc. Let $Q$ be the clockwise endpoint of this arc (which has length $L$ and counterclockwise endpoint $P$). The conditional distribution of $Q$ given $P$ is uniform on the circumference. Thus $P$ and $Q$ are independent.

Clearly the uniform distribution is the only distribution for the arc lengths which makes the endpoints $P$ and $Q$ independent. Denote by $[s,t]$ the arc whose counterclockwise and clockwise endpoints are $s$ and $t$ respectively. Since $P$ and $Q$ are i.i.d. it follows that $[P,Q]$ and $[Q,P]$ are identically distributed. $[P,Q]$ and $[Q,P]$ are complementary arcs: they are disjoint and their union is the entire circumference.

Now consider tossing $n$ independent arcs with uniformly distributed arc lengths. Using the previous observations and notation these $n$ arcs can be written as $[X_1,Y_1],\ldots,[X_n,Y_n]$ where $X_1,\ldots,X_n,Y_1,\ldots,Y_n$ are independent and distributed uniformly on the circumference. Choose $u$ uniformly at random from the elements of $\{X_1,\ldots,X_n,Y_1,\ldots,Y_n\}$. Let $Z_1,Z_2,\ldots,Z_{2n}$ be the clockwise ordering of the points $X_1,\ldots,X_n,Y_1,\ldots,Y_n$ with $Z_1 = u$. Define $\underset{\sim}{Z} = (Z_1,\ldots,Z_{2n})$. Let $\underset{\sim}{\xi} = ((\xi_1,\eta_1),(\xi_2,\eta_2),\ldots,(\xi_n,\eta_n))$ be defined by $(\xi_k,\eta_k) = (i,j)$ if and only if $i < j$ and $\{Z_i,Z_j\} = \{X_k,Y_k\}$. Now define the orientations $\underset{\sim}{\lambda} = (\lambda_1,\ldots,\lambda_n)$ by

$$\lambda_k = \begin{cases} 1 & \text{if } [Z_{\xi_k},Z_{\eta_k}] = [X_k,Y_k] \,, \\ -1 & \text{if } [Z_{\eta_k},Z_{\xi_k}] = [X_k,Y_k] \,. \end{cases}$$

Thus $\underset{\sim}{Z}$ determines the $2n$ endpoints, $\underset{\sim}{\xi}$ partitions the endpoints into $n$ pairs, and $\underset{\sim}{\lambda}$ orients each of these $n$ pairs to obtain $n$ arcs (each pair of points $\{s,t\}$ has two orientations which give the complementary arcs $[s,t]$ and $[t,s]$). Since $X_1,\ldots,X_n,Y_1,\ldots,Y_n$ are i.i.d. uniform, $\underset{\sim}{Z}$, $\underset{\sim}{\lambda}$ and $\underset{\sim}{\xi}$ are independent with $\underset{\sim}{\xi}$ uniformly distributed on its set of $(2n)!/2^n$ possible values and $\underset{\sim}{\lambda}$ uniformly distributed on its set of $2^n$ possible values.

With the usual convention $Z_{2n+1} \equiv Z_1$ define $S_k$ to be the length of $[Z_k, Z_{k+1}]$. Let $\underset{\sim}{S} = (S_1, \ldots, S_{2n})$. $\underset{\sim}{S}$ consists of the spacings between $2n$ points chosen uniformly and independently on the circumference. For

$$1 \leq i \leq 2n \quad \text{let} \quad M_i = |\{k: [Z_i, Z_{i+1}] \subset [X_k, Y_k]\}|$$

so that $M_i$ is the number of arcs $[X_k, Y_k]$ which cover $[Z_i, Z_{i+1}]$. Define the covering vector $\underset{\sim}{M} = (M_1, \ldots, M_{2n})$. $\underset{\sim}{M}$ is determined by $\underset{\sim}{\xi}$ and $\underset{\sim}{\lambda}$ since

$$M_i = \sum_{k=1}^{n} J_{ik} \quad \text{where}$$

(*)

$$J_{ik} = I_{\{\lambda_k=1, \xi_k \leq i < \eta_k\}} + I_{\{\lambda_k=-1\}}(1 - I_{\{\xi_k \leq i < \eta_k\}}) .$$

The $\lambda_k$ are i.i.d. and independent of $\underset{\sim}{\xi}$ so that $J_{i1}, J_{i2}, \ldots, J_{in}$ are conditionally independent given $\underset{\sim}{\xi}$ with

$$P\{J_{ik} = 1 | \underset{\sim}{\xi}\} = P\{J_{ik} = 0 | \underset{\sim}{\xi}\} = \frac{1}{2} .$$

Thus the distribution of $M_i$ is Binomial $(n, \frac{1}{2})$ .

Define

$$G_k = \sum_{i=1}^{2n} I_{\{M_i = k\}}$$

and

$$V_k = \sum_{i=1}^{2n} S_i I_{\{M_i = k\}} .$$

$G_k$ is the number of segments $[Z_i, Z_{i+1}]$ which are covered $k$ times.

$V_k$ is the length of that portion of the circumference which is covered $k$ times. To relate this with the notation of the previous sections note that $G = G_0$ and $V = V_0$. From the definition it follows immediately that

$$E\, G_k = 2n\, P\{M_i = k\} = \frac{2n\binom{n}{k}}{2^n}\,.$$

Since $\underset{\sim}{S}$ and $(\underset{\sim}{\xi},\underset{\sim}{\lambda})$ are independent it is immediate that

$$E\, V_k = 2n(E\, S_i)(P\{M_i = k\}) = \frac{\binom{n}{k}}{2^n}\,.$$

The random variables $G_k$ and $V_k$ obey some obvious constraints. $|M_i - M_{i+1}| = 1$ for all $i$ where $M_{2n+1} \equiv M_1$. Thus $M_i$ is an even number if and only if $M_{i+1}$ is odd which implies $\sum_{k\text{ even}} G_k = \sum_{k\text{ odd}} G_k = n$. Also $\sum_{k=0}^{n} V_k = 1$.

$\underset{\sim}{M}$ is determined by $\underset{\sim}{\xi}$ and $\underset{\sim}{\lambda}$ as shown previously. Write this dependence as $\underset{\sim}{M} = f(\underset{\sim}{\xi},\underset{\sim}{\lambda})$. Define $\underset{\sim}{M}' = f(\underset{\sim}{\xi},-\underset{\sim}{\lambda})$ where $-\underset{\sim}{\lambda} = (-\lambda_1,\ldots,-\lambda_n)$. $-\underset{\sim}{\lambda}$ and $\underset{\sim}{\lambda}$ are identically distributed and $\underset{\sim}{\xi}$ and $\underset{\sim}{\lambda}$ are mutually independent. Thus $\underset{\sim}{M}'$ and $\underset{\sim}{M}$ are identically distributed. $\underset{\sim}{M}'$ is the covering vector that would be obtained using the complementary arcs $[Y_1,X_1],\ldots,[Y_n,X_n]$. From (*) it follows that $(M_1',\ldots,M_{2n}') = (n-M_1,\ldots,n-M_{2n})$. Thus $(n-M_1,\ldots,n-M_{2n})$ and $(M_1,\ldots,M_{2n})$ are identically distributed. From this it follows immediately that $G_i$ and $G_{n-i}$ are identically distributed for all $i$. Similarly $V_i$ and $V_{n-i}$ are identically distributed.

---

The ordered pair $(\underset{\sim}{\xi},\underset{\sim}{\lambda})$ is uniformly distributed over the set of $(2n)!$ possible values. Thus any random variable (such as $G_i$) which is a function of $\underset{\sim}{\xi}$ and $\underset{\sim}{\lambda}$ can be investigated by purely combinatorial methods.

$\underset{\sim}{S}$ is independent of $\underset{\sim}{\xi}$ and $\underset{\sim}{\lambda}$ so that $\mathcal{L}(V_k|G_k=j) = \mathcal{L}(S_1+\cdots+S_j)$. Thus

$$E(V_k^p|G_k=j) = E(S_1+\cdots+S_j)^p = \frac{\binom{j+p-1}{p}}{\binom{2n+p-1}{p}}$$

and consequently

$$E\, V_k^p = E\{E(V_k^p|G_k)\} = \binom{2n+p-1}{p}^{-1} E\binom{G_k+p-1}{p} .$$

Using an elementary binomial identity this becomes

$$(4.1) \qquad E\, V_k^p = \binom{2n+p-1}{p}^{-1} E\binom{G_k+p-1}{p} = \binom{2n+p-1}{p}^{-1} \sum_{j=1}^{p} \binom{p-1}{j-1} E\binom{G_k}{j} .$$

This expression allows the moments of $V_k$ to be computed combinatorially. Formula (4.1) can itself be proved by combinatorial arguments (which use the same basic idea as the third proof of (2.5)). However, even though most quantities of interest can be computed combinatorially, analytic arguments are generally preferred because they are usually less tedious and notationally simpler.

All the moments of $G_0$ and $V_0$ are obtained by straightforward evaluation of the expressions (3.1) and (3.2) (just use the multinomial expansion followed by the Dirichlet integral). One way of writing these moments which is useful for purposes of comparison is given below.

$$\text{(4.2)} \qquad E\binom{G_0}{q} = \frac{2n}{q} \cdot \frac{n!}{2^n(2n)!} \sum_{\sigma} \prod_{i=1}^{q} \frac{(2\sigma_i)!}{\sigma_i!}$$

with summation over all q-tuples $\sigma = (\sigma_1,\ldots,\sigma_q)$ satisfying $\Sigma\, \sigma_i = n$ and $\sigma_i > 0$ for all $i$.

$$\text{(4.3)} \qquad E\, V_0^p = \frac{2n}{p} \left(\frac{n!}{2^n(2n)!}\right)\binom{2n+p-1}{p}^{-1} \sum_{\sigma} \prod_{i=1}^{p} \frac{(2\sigma_i)!}{\sigma_i!} \;,$$

with summation over all p-tuples $\sigma = (\sigma_1,\ldots,\sigma_p)$ satisfying $\Sigma\, \sigma_i = n$ and $\sigma_i \geq 0$ for all $i$. The crucial difference between these two expressions is that $\sigma_i > 0$ in (4.2) and $\sigma_i \geq 0$ in (4.3). Using (4.2) and (4.3) allows one to verify by inspection the validity of (4.1) in the case $k = 0$.

The first moments of $G_k$ and $V_k$ were given in the previous subsection. The second moments of $V_k$ and $G_k$ for arbitrary $k$ are now obtained. Choose points $Q_1$ and $Q_2$ distributed uniformly and independently on the circumference and independent of $X_1,\ldots,X_n$, $Y_1,\ldots,Y_n$. Define $W_i = |\{k\colon Q_i \in [X_k,Y_k]\}|$. $W_i$ is the number of arcs which contain $Q_i$. Following the derivation of (2.7) gives $E\, V_t^2 = P\{W_1 = W_2 = t\}$.

An arc $[X_k,Y_k]$ will be said to separate $Q_1$ and $Q_2$ if it contains one of the points but not the other. Observe that $[X_k,Y_k]$ separates $Q_1$ and $Q_2$ if and only if the complementary arc $[Y_k,X_k]$ separates $Q_1$ and $Q_2$. Let $\theta$ be the number of arcs which separate $Q_1$ and $Q_2$. This may be written as

$$\theta = \sum_{k=1}^{n} \left| I_{\{Q_1 \in [X_k,Y_k]\}} - I_{\{Q_2 \in [X_k,Y_k]\}} \right| \;.$$

The above observation says that $\theta$ does not depend on the orientations $\underset{\sim}{\lambda}$ of the arcs and so must be a function of $Q_1$, $Q_2$, $\underset{\sim}{Z}$ and $\underset{\sim}{\xi}$.

Condition on $Q_1$, $Q_2$, $\underset{\sim}{Z}$ and $\underset{\sim}{\xi}$ so that these quantities may be considered as fixed in the following argument. Let $A$ be the number of arcs which contain $Q_1$ but not $Q_2$. Let $B$ be the number of arcs which contain both $Q_1$ and $Q_2$. Then $W_1 = A+B$ and $W_2 = (\theta-A)+B$. Since the orientations $\lambda_1,\dots,\lambda_n$ are i.i.d. (and independent of $Q_1$, $Q_2$, $\underset{\sim}{Z}$ and $\underset{\sim}{\xi}$) with $P\{\lambda_i=1\} = P\{\lambda_i=-1\} = \frac{1}{2}$, the random variables $A$ and $B$ are independent with $A \sim \text{Binomial}\ (\theta, \frac{1}{2})$ and $B \sim \text{Binomial}\ (n-\theta,\frac{1}{2})$. $W_1 = W_2 = t$ if and only if $A = \frac{\theta}{2}$ and $B = t-\frac{\theta}{2}$. Thus the event $\{W_1 = W_2 = t\}$ is impossible unless $\theta$ is even. Set $\theta = 2k$. Then

$$P\{A=\tfrac{\theta}{2}\} = \frac{\binom{2k}{k}}{2^{2k}} \quad \text{and} \quad P\{B=t-\tfrac{\theta}{2}\} = \frac{\binom{n-2k}{t-k}}{2^{n-2k}} \quad .$$

Multiplying these probabilities gives

$$P\{W_1 = W_2 = t \,|\, Q_1, Q_2, \underset{\sim}{Z}, \underset{\sim}{\xi}\} = \frac{1}{2^n}\binom{2k}{k}\binom{n-2k}{t-k} \quad .$$

This conditional probability depends only on $\theta$ so that

$$P\{W_1 = W_2 = t \,|\, \theta = 2k\} = \frac{1}{2^n}\binom{2k}{k}\binom{n-2k}{t-k} \; .$$

It remains to evaluate $P\{\theta=m\}$. Condition on $Q_1$ and $Q_2$ so that these points are fixed in the following argument. Let $R_1$ be the length of the arc $[Q_1,Q_2]$ and $R_2 = 1-R_1$. Let $I_k$ be the indicator of the event

$\{[X_k,Y_k]$ separates $Q_1$ and $Q_2\}$. Then $\theta = \Sigma_{k=1}^n I_k$. $I_1,\ldots,I_n$ are clearly i.i.d. with $P\{I_k = 1\} = 2R_1R_2$. Thus $\mathcal{L}(\theta|Q_1,Q_2) = \text{Binomial}\ (n, 2R_1R_2)$ and

$$P\{\theta = m|Q_1,Q_2\} = \binom{n}{m}(2R_1R_2)^m(R_1^2+R_2^2)^{n-m}$$

since $1-2R_1R_2 = R_1^2+R_2^2$. Taking expectations yields

$$P\{\theta = m\} = \binom{n}{m}2^m\ E\{R_1^mR_2^m(R_1^2+R_2^2)^{n-m}\}$$

where $R_1$ is uniform on $(0,1)$ and $R_1+R_2 = 1$. Use the binomial expansion followed by the beta integral to evaluate this expression and obtain

$$P\{\theta = m\} = \frac{2^m\binom{n}{m}}{2n+1}\sum_{j=0}^{n-m}\binom{n-m}{j}\binom{2n}{m+2j}^{-1}.$$

Substituting the previous results in the expression

$$P\{W_1 = W_2 = t\} = \sum_k P\{W_1 = W_2 = t|\theta = 2k\}P\{\theta = 2k\}$$

and simplifying the binomial coefficients yields

$$\text{(4.4)} \qquad E\ V_t^2 = \frac{\binom{n}{t}}{2^n(2n+1)}\sum_k 4^k\binom{t}{k}\binom{n-t}{k}\sum_j\binom{n-2k}{j}\binom{2n}{2j+2k}^{-1}$$

with the sums taken over all integers $k$ and $j$ and assuming the usual conventions $\binom{x}{y} = 0$ if $y < 0$ or $x < y$.

From (4.1) it follows that

$$E\, V_t^2 = \binom{2n+1}{2}^{-1} \{E\, G_t + E\binom{G_t}{2}\} \quad .$$

Since $E\, G_t = 2n\binom{n}{t}/2^n$ formula (4.4) also allows immediate calculation of $E\binom{G_t}{2}$.

The expression (4.4) can be rewritten to more closely parallel (4.3):

$$E\, V_t^2 = n\{\frac{n!}{2^n(2n)!}\} \binom{2n+1}{2}^{-1} \sum_{\sigma} \frac{(2\sigma_1)!(2\sigma_2)!}{\sigma_1!\sigma_2!} \sum_k 4^k \binom{n-2k}{t-k}\binom{\sigma_1}{k}\binom{\sigma_2}{k}$$

where the summation is over $\sigma = (\sigma_1,\sigma_2)$ satisfying $\sigma_1+\sigma_2 = n$ and $\sigma_1 \geq 0$, $\sigma_2 \geq 0$. This expression can be modified to obtain a formula for $E\binom{G_t}{2}$ by simply deleting the factor of $\binom{2n+1}{2}^{-1}$ and requiring $\sigma_1 > 0$ and $\sigma_2 > 0$ in the sum.

An expression for $E\binom{G_1}{p}$ will now be given. For $\tau \subset \{1,2,\ldots,2n-1,2n\}$ define $H_\tau = \bigcap_{i\in\tau} \{M_i = 1\}$. Then $\binom{G_1}{p} = \sum_{\tau\in T} I_{H_\tau}$ with $T = \{\tau : |\tau| = p\}$. Thus $E\binom{G_1}{p} = \sum_{\tau\in T} P(H_\tau)$. Since $(\underset{\sim}{\xi},\underset{\sim}{\lambda})$ is uniformly distributed over the set of $(2n)!$ possible values $P(H_\tau) = \frac{1}{(2n)!} |\{(\underset{\sim}{\xi},\underset{\sim}{\lambda}) : M_i = 1 \text{ for } i \in \tau\}|$. Let $\tau = \{\tau_1,\tau_2,\ldots,\tau_p\}$ with $\tau_1 < \tau_2 < \ldots < \tau_p$. Define $\sigma_i = \frac{1}{2}(\tau_{i+1}-\tau_i)$ for $1 \leq i \leq p-1$ and $\sigma_p = n - (\sigma_1 + \cdots + \sigma_{p-1})$. Since $|M_i - M_{i+1}| = 1$ for all $i$ it is clear that $P(H_\tau) = 0$ unless $\sigma_1,\ldots,\sigma_p$ are all positive integers.

Routine but tedious counting arguments show that

$$|\{(\underset{\sim}{\xi},\underset{\sim}{\lambda}) : M_i = 1 \text{ for } i \in \tau\}| = \frac{n!}{2^n} \{(\prod_{i=1}^{p} (2\sigma_i+1)) - (n+1)\} \prod_{i=1}^{p} \frac{(2\sigma_i)!}{\sigma_i!} \quad .$$

To motivate this answer note that

$$|\{(\underset{\sim}{\xi},\underset{\sim}{\lambda}): M_i = 0 \quad \text{for} \quad i\varepsilon\tau\}| = \frac{n!}{2^n} \prod_{i=1}^{p} \frac{(2\sigma_i)!}{\sigma_i!}$$

Now show that every pair $(\underset{\sim}{\xi},\underset{\sim}{\lambda})$ which leads to $M_i = 0$ for $i \varepsilon \tau$ can be modified in $\{(\prod_{i=1}^{p} (2\sigma_i+1)) - (n+1)\}$ different ways to yield pairs $(\underset{\sim}{\xi},\underset{\sim}{\lambda})$ with $M_i = 1$ for $i \varepsilon \tau$.

Dividing by $(2n)!$ and summing over $\tau \varepsilon T$ gives

$$(4.5) \qquad E\binom{G_1}{p} = \frac{2n}{p} \cdot \frac{n!}{2^n(2n)!} \sum_{\sigma} \{(\prod_{i=1}^{p} (2\sigma_i+1))-(n+1)\} \prod_{i=1}^{p} \frac{(2\sigma_i)!}{\sigma_i!}$$

with the summation over all p-tuples $\sigma = (\sigma_1,\ldots,\sigma_p)$ satisfying $\Sigma\sigma_i = n$ and $\sigma_i > 0$ for all $i$. The factor $\frac{2n}{p}$ arises when the sum over the subsets $\tau$ is transformed into a sum over the partitions $\sigma$.

To obtain $E\, V_1^p$ use (4.5) to calculate $E\binom{G_1}{1}, E\binom{G_1}{2},\ldots,E\binom{G_1}{p}$ and then plug these values into expression (4.1). Expression (4.5) can be modified to obtain a formula for $E\, V_1^p$ by throwing in a factor of $\binom{2n+p-1}{p}^{-1}$ and extending the sum to allow $\sigma_i \geq 0$ for all $i$. Expressions for $E\binom{G_k}{p}$ along the lines of (4.5) can be found for arbitrary $k$ and $p$. They are not given here because (in their present form) these expressions are quite cumbersome.

---

Define $\underset{\sim}{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_{2n})$ by $\zeta_i = M_i - M_{i-1}$ where $M_0 \equiv M_{2n}$. An equivalent definition is

$$\zeta_i = \begin{cases} 1 & \text{if } Z_i \in \{X_1, \ldots, X_n\} , \\ -1 & \text{if } Z_i \in \{Y_1, \ldots, Y_n\} . \end{cases}$$

Define $S_0, S_1, \ldots, S_{2n}$ by $S_0 = 0$ and $S_k = \sum_{i=1}^{k} \zeta_i = M_i - M_0$ for $k > 0$. This notation differs from that in the preceding subsections. Let $B = \{\underset{\sim}{\zeta} : |\zeta_i| = 1 \text{ for all } i \text{ and } \Sigma\, \zeta_i = 0\}$ and $B^+ = \{\underset{\sim}{\zeta} : \zeta \in B \text{ and } S_i \geq 0 \text{ for all } i\}$. $B$ is the set of bridges and $B^+$ is the set of positive bridges.

It is clear that $\underset{\sim}{\zeta}$ is uniformly distributed over the $\binom{2n}{n}$ elements of $B$. Let $1 \leq \tau_1 < \tau_2 < \cdots < \tau_p \leq 2n$. Using the uniform distribution of $\underset{\sim}{\zeta}$ and simple counting arguments one can show that

$$P\{M_{\tau_1} = M_{\tau_2} = \cdots = M_{\tau_p}\} = P\{S_{\tau_1} = S_{\tau_2} = \cdots = S_{\tau_p}\} = \binom{2n}{n}^{-1} \prod_{i=1}^{p} \binom{2\sigma_i}{\sigma_i} ,$$

where $\sigma_i = \frac{1}{2}(\tau_{i+1} - \tau_i)$ for $1 \leq i \leq p-1$ and $\sigma_p = n - (\sigma_1 + \cdots + \sigma_{p-1})$. Let $\tau = (\tau_1, \ldots, \tau_p)$ and define $T = \{\tau : 1 \leq \tau_1 < \tau_2 < \cdots < \tau_p \leq 2n\}$. Since

$$\sum_{k=0}^{n} \binom{G_k}{p} = \sum_{\tau \in T} I_{\{M_{\tau_1} = M_{\tau_2} = \cdots = M_{\tau_p}\}} ,$$

it follows that

$$\sum_{k=0}^{n} E\binom{G_k}{p} = \sum_{\tau \in T} P\{M_{\tau_1} = M_{\tau_2} = \cdots = M_{\tau_p}\}$$

which can be rewritten as

$$(4.6) \qquad \sum_{k=0}^{n} E\binom{G_k}{p} = \frac{2n}{p} \cdot \binom{2n}{n}^{-1} \sum_{\sigma} \prod_{i=1}^{p} \binom{2\sigma_i}{\sigma_i}$$

with the summation over $\sigma = (\sigma_1, \ldots, \sigma_p)$ satisfying $\Sigma \sigma_i = n$ and $\sigma_i > 0$ for all $i$. Using (4.1) gives the companion formula

$$(4.7) \qquad \sum_{k=0}^{n} E\, V_k^p = \frac{2n}{p} \cdot \binom{2n+p-1}{p}^{-1} \binom{2n}{n}^{-1} \sum_{\sigma} \prod_{i=1}^{p} \binom{2\sigma_i}{\sigma_i}$$

with the summation over p-tuples $\sigma = (\sigma_1, \ldots, \sigma_p)$ satisfying $\Sigma \sigma_i = n$ and $\sigma_i \geq 0$ for all $i$. This formula gives a check on (4.4).

$M_k = S_k + M_0 \geq 0$ for all $k$. Thus $M_0 = 0$ implies $\underset{\sim}{\zeta} \in B^+$. $\underset{\sim}{\zeta}$ consists of $n$ positive and $n$ negative coordinates. Each positive coordinate is paired to a negative coordinate by an arc. This pairing can be done in $n!$ equally likely ways. If $M_0 = 0$, then each $j$ with $\zeta_j = -1$ must be paired with a $k$ satisfying $k < j$ and $\zeta_k = 1$. This can be done in $W(\underset{\sim}{\zeta})$ different ways where

$$W(\underset{\sim}{\zeta}) = \prod_{i:\zeta_{i+1} = -1} S_i \ .$$

Therefore

$$P\{M_0 = 0 \,|\, \underset{\sim}{\zeta} = \underset{\sim}{t}\} = \begin{cases} \dfrac{W(\underset{\sim}{t})}{n!} & \text{for } \underset{\sim}{t} \in B^+ , \\ 0 & \text{otherwise} . \end{cases}$$

$W(\underset{\sim}{\zeta})$ can be written in a more symmetrical fashion. Define $L = \{i: S_{i-1} > S_i \text{ and } S_i < S_{i+1}\}$ and $H = \{i: S_{i-1} < S_i \text{ and } S_i > S_{i+1}\}$ where for notational convenience $S_i = 0$ for $i \leq 0$ and $i \geq 2n$. $L$ is the set of relative minima (low spots) and $H$ is the set of relative maxima (high spots). Then

$$W(\underset{\sim}{\zeta}) = \frac{\prod_{i \in H} (S_i!)}{\prod_{i \in L} (S_i!)} .$$

Since $P\{\underset{\sim}{\zeta} = \underset{\sim}{t}\} = \binom{2n}{n}^{-1}$ for all $\underset{\sim}{t}$, summing $P\{M_0 = 0 | \underset{\sim}{\zeta} = \underset{\sim}{t}\}$ over $\underset{\sim}{t}$ yields

$$P\{M_0 = 0\} = \frac{1}{2^n} = \frac{n!}{(2n)!} \sum_{\underset{\sim}{t} \in B^+} W(\underset{\sim}{t}) .$$

This gives the amusing identity

$$\sum_{\underset{\sim}{t} \in B^+} W(\underset{\sim}{t}) = \frac{(2n)!}{n! 2^n} = 1 \times 3 \times 5 \times \cdots \times (2n-1) . \tag{4.8}$$

It is easy to show that $B^+$ contains $\frac{1}{n+1} \binom{2n}{n}$ elements. Assume $\underset{\sim}{T}$ is uniformly distributed over $B^+$. Then (4.8) can be restated as

$$E\, W(\underset{\sim}{T}) = \frac{(n+1)!}{2^n} .$$

## 5. $V_k$ and $G_k$ for Arbitrary Arc Length Distributions.

This section deals with the quantities $V_k$ and $G_k$ introduced in the preceding section. The definitions are repeated for convenience. $n$ arcs are placed at random on a circle of unit circumference. $V_k$ is the measure of that part of the circumference covered exactly $k$ times. The $2n$ points which are the endpoints of the $n$ random arcs divide the circumference into $2n$ segments. $G_k$ is the number of segments covered exactly $k$ times. The previous section dealt only with the case in which the arc lengths had a uniform distribution. Arbitrary distributions $F$ for the arc length are now considered.

The expectation of $V_k$ is easily calculated. Let $Q$ be a point chosen uniformly on the circumference and independent of the $n$ random arcs. Let $W$ be the number of arcs covering $Q$. Arguing as in section 3 gives $EV_k = P\{W=k\}$. $W = \Sigma_j I_j$ where $I_j$ is the indicator of the event that the $j^{th}$ arc covers $Q$. The $I_j$ are clearly independent and identically distributed with $EI_j = \mu \equiv \int_0^1 (1-F(x))dx$. $\mu$ is the expected arc length. Thus $W$ is Binomial $(n,\mu)$ so that

$$EV_k = \binom{n}{k}\mu^k(1-\mu)^{n-k} . \tag{5.1}$$

The second moments $EV_jV_k$ are more complicated. Label the $n$ random arcs by the integers $1,2,\ldots,n$. Let $\pi \subset \{1,2,\ldots,n\}$. Define $U_\pi$ to be the length of that part of the circumference which is not covered by any of the arcs in $\pi$. $U_\pi$ is the measure of the vacancy which results after the arcs in $\pi$ are tossed on the circumference. For $1 \leq j \leq n$ define the functions

$$f_j(z) = \begin{cases} 0 & \text{if } z \text{ is covered by the } j^{th} \text{ arc,} \\ 1 & \text{otherwise .} \end{cases}$$

Then $U_\pi = \int_0^1 \prod_{j\epsilon\pi} f_j(z)dz$. Upon expanding $\prod_{k\notin\pi}(1-f_k(z))$ one obtains

$$\sum_{|\pi|=p} \prod_{j\epsilon\pi} f_j(z) \prod_{k\notin\pi}(1-f_k(z)) = \sum_{|\pi|\geq p}(-1)^{|\pi|-p}\binom{|\pi|}{p}\prod_{j\epsilon\pi} f_j(z) .$$

Integrating both sides and noting that

$$V_{n-p} = \sum_{|\pi|=p}\int_0^1 \prod_{j\epsilon\pi} f_j(z)\prod_{k\notin\pi}(1-f_k(z))dz$$

yields

$$V_{n-p} = \sum_{|\pi|\geq p}(-1)^{|\pi|-p}\binom{|\pi|}{p}U_\pi . \tag{5.2}$$

The moments $EU_\pi U_\sigma$ may be evaluated by the methods of sections 2 and 3. Let $Q_1$ and $Q_2$ be chosen uniformly and independently on the circumference (and independent of the $n$ random arcs). Then $EU_\pi U_\sigma = P(B)$ where $B$ is the event that $Q_1$ is not covered by any of the arcs in $\pi$ and $Q_2$ is not covered by any of the arcs in $\sigma$. Let $B_{ij}$ be the event that $Q_i$ is not covered by the $j^{th}$ arc. Then

$$(*) \quad B = (\prod_{j\epsilon\pi} B_{1j})(\prod_{k\epsilon\sigma} B_{2k}) = (\prod_{j\epsilon\pi-\sigma} B_{1j})(\prod_{k\epsilon\sigma-\pi} B_{2k})(\prod_{\ell\epsilon\pi\cap\sigma} B_{1\ell}B_{2\ell})$$

where the product notation is used to denote intersection. The factor events in (*) are conditionally independent given $Q_1$ and $Q_2$. The arguments of section 3 yield

$$P(B_{1j}|Q_1,Q_2) = P(B_{2k}|Q_1,Q_2) = g(1)$$

and

$$P(B_{1\ell}B_{2\ell}|Q_1,Q_2) = g(S_1) + g(S_2) ,$$

where $g(s) \equiv \int_0^s F(x)dx$ and $S_1,S_2$ are the spacings between the points $Q_1,Q_2$. Let $\alpha = g(1)$. Note that $\alpha = \int_0^1 F(x)dx = 1-\mu$. Then using the conditional independence gives

$$P(B|Q_1,Q_2) = \alpha^{|\sigma-\pi|+|\pi-\sigma|}\{g(S_1)+g(S_2)\}^{|\sigma\cap\pi|} .$$

Thus

$$(5.3) \qquad EU_\pi U_\sigma = \alpha^{|\sigma-\pi|+|\pi-\sigma|}E'g(S_1)+g(S_2)\}^{|\sigma\cap\pi|} ,$$

where $S_1$ and $S_2$ are the spacings between two points chosen uniformly at random on the circumference. Define $M(r) = E\{g(S_1)+g(S_2)\}^r$. Equation (3.2) says that $M(r) = E_rV^2$ where $r$ is the number of random arcs.

Now use (5.2) and (5.3) to calculate

$$EV_{n-p}V_{n-q} = \sum_\pi\sum_\sigma (-1)^{|\pi|+|\sigma|-p-q}\binom{|\pi|}{p}\binom{|\sigma|}{q} \cdot \alpha^{|\pi-\sigma|+|\sigma-\pi|}M(|\sigma\cap\pi|)$$

$$= (-1)^{p+q}\sum_\tau M(|\tau|)\sum_{\sigma,\pi}(-\alpha)^{|\sigma\cup\pi|}\binom{|\tau\cup\pi|}{p}\binom{|\tau\cup\sigma|}{q} .$$

The transformation $\tau' = \sigma \cap \pi$, $\pi' = \pi - \sigma$ and $\sigma' = \sigma - \pi$ is used to obtain the last expression. Thus the summation in this last expression is over all mutually disjoint subsets $\pi, \tau$ and $\sigma$. As usual assume that $\binom{x}{y} = 0$ if $x < y$. Counting the choices of $\tau, \pi$ and $\sigma$ with $|\tau| = k$, $|\pi| = i$ and $|\sigma| = j$ yields

$$= (-1)^{p+q} \sum_k \binom{n}{k} M(k) \sum_{i,j} (-\alpha)^{i+j} \binom{n-k}{i,j} \binom{k+i}{p} \binom{k+j}{q}$$

where $\binom{a}{b,c}$ denotes the usual multinomial coeffieint

$$\binom{a}{b,c} = \begin{cases} \dfrac{a!}{b!c!(a-b-c)!} & \text{if } b \geq 0,\ c \geq 0 \text{ and } b+c \leq a\ , \\ 0 & \text{otherwise}\ . \end{cases}$$

This can be rewritten as

$$(5.4) \qquad EV_{n-p}V_{n-q} = (-1)^{p+q} \sum_k M(k) \sum_\ell \binom{n}{k,\ell} (-\alpha)^\ell \sum_i \binom{\ell}{i} \binom{k+i}{p} \binom{k+\ell-i}{q}$$

where the summation is over all integers $k, \ell, i$ when the usual conventions for multinomial coefficients are followed. Examination of the sum shows that the coefficient of $M(k)$ is zero if $k < p+q-n$.

A simple generating function argument gives

$$\sum_i \binom{\ell}{i} \binom{k+i}{p} \binom{k+\ell-i}{q} = \sum_{i,j} 2^{\ell-i-j} \binom{\ell}{i,j} \binom{k}{p-i} \binom{k}{q-j}\ .$$

This alternative form may be more convenient if p, q or k is small.

Some special cases of (5.4) are

$$EV_1^2 = n^2M(n) + n(1-2n\alpha)M(n-1) + n(n-1)\alpha^2M(n-2)$$

and

$$EV_oV_p = (-1)^p\binom{n}{p}\sum_k \binom{p}{k}(-\alpha)^k M(n-k) .$$

Let the random variable L be distributed according to F. Define $\tilde{F}$ to be the c.d.f. of $(1-L)$. Thus $\tilde{F}(x) = 1-F(1-x)$. The trivial observation that a segment of the circumference is covered by exactly p arcs if and only if it is not covered by exactly n-p arcs yields the following result.

$$\mathcal{L}_F(V_0,V_1,\ldots,V_n) = \mathcal{L}_{\tilde{F}}(V_n,V_{n-1},\ldots,V_0)$$

where $\mathcal{L}_H$ denotes that the arc lengths are distributed according to H. In particular $E_F V_p V_q = E_{\tilde{F}} V_{n-p} V_{n-q}$. Applying this result to the above special cases of (5.4) immediately gives expressions for $EV_{n-1}^2$ and $EV_nV_p$.

General expressions may in principle be obtained for the higher moments of $V_k$. For example the same arguments used to obtain (5.3) can be used to show that

$$(*) \qquad E\,U_{\sigma_1}U_{\sigma_2}U_{\sigma_3} = E\{g(1)\}^{k_1+k_2+k_3}$$

$$\{g(S_1+S_2)+g(S_3)\}^{k_{12}}\{g(S_1+S_3)+g(S_2)\}^{k_{13}}$$

$$\{g(S_2+S_3)+g(S_1)\}^{k_{23}}\{g(S_1)+g(S_2)+g(S_3)\}^{k_{123}} .$$

Here $S_1$, $S_2$ and $S_3$ are the spacings between three points chosen randomly on the circumference and for $\tau \subset \{1,2,3\}$ define $k_\tau = |\prod_{i\in\tau} \sigma_i \prod_{j\notin\tau} \tilde{\sigma}_j|$ with products denoting intersection and "$\sim$" denoting the complement. Using (*) and (5.2) the moments $E\, V_p V_q V_r$ can be easily calculated. However the expressions for these moments will not be given as they are exceedingly cumbersome.

---

The $n$ random arcs divide the circumference into $2n$ segments. Each arc or segment has an initial point (the point which is furthest counterclockwise) and a final point (the point which is furthest clockwise). For $0 \le k \le n-1$ let $C_k$ be the number of arcs whose initial points are covered exactly $k$ times and $\tilde{C}_k$ be the number of arcs whose final points are covered exactly $k$ times.

By associating each segment with its initial point one obtains $G_k = C_{k-1} + \tilde{C}_k$ for $0 \le k \le n$. Associating each segment with its final point yields $G_k = C_k + \tilde{C}_{k-1}$. It is understood that $C_{-1} = \tilde{C}_{-1} = C_n = \tilde{C}_n = 0$. Equating the two expressions for $G_k$ gives $\tilde{C}_k - \tilde{C}_{k-1} = C_k - C_{k-1}$ for $0 \le k \le n$. This implies $C_k = \tilde{C}_k$ for all $k$. Thus for all $k$

$$G_k = C_k + C_{k-1}\ . \tag{5.5}$$

This expression is useful because $C_k$ is easier to handle than $G_k$.

The calculation of $EC_k$ is immediate. Let $I_{jk}$ be the indicator of the event that the initial point of the $j^{th}$ arc is covered exactly $k$ times. Then $C_k = \Sigma_j I_{jk}$ so that $EC_k = nEI_{1k} = n\cdot\text{Prob}\{P_1$ is covered $k$ times$\}$ where $P_1$ is the initial point of the first arc. Since $P_1$

is uniform and independent of the $n-1$ arcs labeled by $2,3,\ldots,n$ it is clear that $\text{Prob}\{P_1 \text{ is covered } k \text{ times}\} = E_{n-1}V_k$ $= \binom{n-1}{k}\mu^k(1-\mu)^{n-1-k}$ by (5.1). Thus

$$E\,C_k = n\binom{n-1}{k}\mu^k(1-\mu)^{n-1-k} . \tag{5.6}$$

$EG_k$ is now obtained using (5.5) and (5.6).

The second moments $EC_pC_q$ can be calculated in a straightforward manner similar to the derivation of (5.4). The argument is sketched below. $C_pC_q = \Sigma_i\Sigma_j\, I_{ip}I_{jq}$ so that $EC_pC_q = n(n-1)EI_{1p}I_{2q} + nEI_{1p}I_{1q}$. Now $nEI_{1p}I_{1q} = n\delta(p-q)EI_{1p} = \delta(p-q)EC_p = \delta(p-q)n\binom{n-1}{p}\mu^p(1-\mu)^{n-1-p}$ where

$$\delta(x) = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{otherwise} . \end{cases}$$

$P_i$ is the initial point of the $i^{th}$ arc. Let $J_1$ and $J_2$ be the indicators of the events $\{P_1$ is not covered by the second arc$\}$ and $\{P_2$ is not covered by the first arc$\}$ respectively. Let $H(i,k)$ be the indicator of the event that $P_i$ is covered by exactly $k$ of the arcs with labels in $\{3,4,\ldots,n\}$. With this notation

$$I_{1p} = J_1H(1,p) + (1-J_1)H(1,p-1)$$

and

$$I_{2q} = J_2H(2,q) + (1-J_2)H(2,q-1)$$

so that

$$(5.7)\qquad I_{1p}I_{2q} = H(1,p-1)H(2,q-1)$$

$$+ J_1H(1,p)H(2,q-1) + J_2H(1,p-1)H(2,q)$$

$$- (J_1+J_2)H(1,p-1)H(2,q-1)$$

$$+ J_1J_2\{H(1,p)H(2,q) - H(1,p)H(2,q-1)$$

$$- H(1,p-1)H(2,q) + H(1,p-1)H(2,q-1)\} .$$

Arguing as in (5.2) will give

$$(5.8)\qquad H(i,n-2-s) = \sum_{\pi} (-1)^{|\pi|-s} \binom{|\pi|}{s} U(i,\pi)$$

where the sum is over $\pi \subset \{3,4,\ldots,n\}$ and $U(i,\pi)$ is the indicator of the event that $P_i$ is not covered by any of the arcs in $\pi$. The analogue of (5.3) in this situation is easily seen to be

$$(5.9)\qquad EJ_1^{\varepsilon_1}J_2^{\varepsilon_2}H(1,\pi)H(2,\sigma)$$

$$= \alpha^{|\sigma-\pi| + |\pi-\sigma|} E\, F(S_1)^{\varepsilon_1}F(S_2)^{\varepsilon_2}\{g(S_1)+g(S_2)\}^{|\sigma\cap\pi|}$$

where $\varepsilon_i = 0$ or $1$ and $S_1,S_2$ are as in (5.3). Now expand (5.7) using (5.8) and evaluate the expectation of each term using (5.9) to obtain the desired result.

Define $M_j(k) = E\, F(S_1)^{\varepsilon_1}F(S_2)^{\varepsilon_2}\{g(S_1)+g(S_2)\}^k$ where $j = \varepsilon_1+\varepsilon_2$. The quantities $M_j(k)$ are clearly related to the joint moments in equation (3.3). Let $M_j$ denote the sequence $M_j(0),M_j(1),M_j(2),\ldots$ . Now define

$$R(n,p,q,k) = (-1)^{p+q} \sum_{\ell} \binom{n}{k,\ell}(-\alpha)^{\ell} \quad \sum_{i} \binom{\ell}{i}\binom{k+i}{n-p}\binom{k+\ell-i}{n-q}$$

and let $R(n,p,q)$ denote the sequence $R(n,p,q,0)$, $R(n,p,q,1)$, $R(n,p,q,2),\ldots$ . With this notation (5.4) may be written as an inner product $EV_pV_q = \langle M_0|R(n,p,q)\rangle$.

This notation has been developed in order to allow $EC_pC_q$ to be written in the following reasonably compact form:

$$\text{(5.10)} \qquad EC_pC_q = n(n-1)\{\langle M_0|R(n-2,p-1,q-1)\rangle$$
$$+\langle M_1|R(n-2,p,q-1) + R(n-2,p-1,q) - 2R(n-2,p-1,q-1)\rangle$$
$$+\langle M_2|R(n-2,p,q) - R(n-2,p-1,q) - R(n-2,p,q-1) + R(n-2,p-1,q-1)\rangle\}$$
$$+ \delta(p-q)n\binom{n-1}{p}\mu^p(1-\mu)^{n-1-p} .$$

Observe that the inner product terms precisely parallel the terms in (5.7). Using (5.10) and (5.5) one obtains the moments $EG_pG_q$.

The moments $EV_pC_q$ can also be expressed in a form resembling (5.4) and (5.10). The derivation is sketched below. Let $Q$ be a point chosen uniformly at random on the circumference and independent of the $n$ random arcs. Let $A_p$ be the indicator of the event $\{Q$ is covered by exactly $p$ arcs$\}$ and $I_{iq}$ be the indicator of the event that the initial point of the $i^{th}$ arc is covered by exactly $q$ arcs. Then $EV_pC_q = EV_p\Sigma_i I_{iq} = nEV_pI_{1q} = nEA_pI_{1q}$. Now let $J$ be the indicator of the event that $Q$ is not covered by the first arc and $H_k$ be the indicator of the event that $Q$ is covered by exactly $k$ arcs with labels in $\{2,3,\ldots,n\}$. Then

$$A_p I_{1q} = JH_p I_{1q} + (1-J)H_{p-1} I_{1q}$$

$$(*) \qquad = H_{p-1} I_{1q} + J(H_p I_{1q} - H_{p-1} I_{1q}) \; .$$

Let $M_j(k) = EF(S_1)^j \{g(S_1) + g(S_2)\}^k$ for $j = 0$ or $1$ and let the sequences $M_j$ and $R(n,p,q)$ be as in (5.10). From (*) one can immediately read off the desired expression:

$$(5.11) \qquad EV_p C_q = n\{\langle M_0 | R(n-1,p-1,q)\rangle + \langle M_1 | R(n-1,p,q) - R(n-1,p-1,q\rangle\}.$$

Using (5.5) and (5.11) one obtains the moments $EV_p G_q$.

## 6. A Recursion for the Joint Distribution of Various Sums of Spacings.

This section gives a result concerning the spacings $S_1, S_2, \ldots, S_n$ between $n$ points chosen uniformly on the circumference. For any $\sigma \subset \{1,2,\ldots,n\}$ define

$$S(\sigma) = \sum_{i \in \sigma} S_i .$$

Let $\tau_k = \{1,2,\ldots,k\}$ and $\tau_k + i = \{1+i, 2+i, \ldots, k+i\}$ where addition is modulo $n$ in the sense that $n+1 = 1$, etc. Define

$$H_k(a) = \sum_{i=1}^{n} I_{\{S(\tau_k + i) > a\}} .$$

This quantity arises naturally when considering the random placement of arcs of length = a and in the construction of some tests of uniformity.

In order to compute the second moments $E\, H_j(a) H_k(b)$ one must be able to calculate quantities like $P\{S(\sigma) > a,\ S(\tau) > b\}$. This probability can be written down almost by inspection when $\sigma \subset \tau$ or when $\sigma \cap \tau = \phi$. A recursion formula is now developed to handle the case when $|\sigma \cap \tau| > 0$.

All events are subsets of the simplex $\{(x_1, x_2, \ldots, x_n) \colon \Sigma_j x_j = 1$ and $x_j > 0$ for all $j\}$. Let $T\colon R^n \to R^n$ be a bijection and denote $T(x_1, x_2, \ldots, x_n) = (x_1', x_2', \ldots, x_n')$. Assume that $T$ has Jacobian equal to 1 and satisfies $x_1 + \cdots + x_n = x_1' + \cdots + x_n'$ for all points $\underset{\sim}{x} \in R^n$. Assume also that $A$ and $B = T(A)$ are both subsets of the simplex. The spacings $(S_1, S_2, \ldots, S_n)$ are uniformly distributed on the simplex. Therefore $P(A) = P(B)$. This simple fact is the basis of the following argument.

Choose $\sigma$ and $\tau$ such that $(\sigma \cup \tau) \cap \{1,2,3\} = \phi$. $\square$ will serve to denote a fixed collection of inequalities of the form $S(\pi) > c$ with $\pi$ satisfying either $\pi \cap \{1,2,3\} = \phi$ or $\pi \supset \{1,2,3\}$. Thus $\{\square\} = \bigcap_{i=1}^{t} \{S(\pi_i) > c_i\}$ for some $t$ and $\pi_1,\ldots,\pi_t$ satisfying the stated condition. Clearly

$$(*) \qquad P\{S_1+S_2+S(\sigma) > a,\ S_1+S_3+S(\tau) > b,\ \square\}$$

$$= P\{S_1+S(\sigma) > a,\ S_1+S_3+S(\tau) > b,\ \square\}$$

$$+ P\{S_1+S(\sigma) \le a < S_1+S_2+S(\sigma),\ S_1+S_3+S(^-) > b,\ \square\}.$$

The transformation

$$S_1' = a-S_1-S(\sigma)\ ,$$

$$S_2' = S_1+S_2+S(\sigma)-a\ ,$$

$$S_3' = S_1+S_3\ , \text{ and}$$

$$S_i' = S_i \text{ for } i > 3$$

satisfies $S_1 + \cdots + S_n = S_1' + \cdots + S_n'$ and has Jacobian $= 1$. The inequalities in $\square$ are unchanged by the transformation. The simplex conditions $\{\Sigma_j S_j = 1,\ S_1 > 0,\ldots,S_n > 0\}$ are always implied when writing an event. With this in mind it is easily shown that the event $\{S_1+S(\sigma) \le a < S_1+S_2+(\sigma),\ S_1+S_3+S(\tau) > b,\ \square\}$ transforms to (after deleting the superscript primes) the event

$$\{S_1+S(\sigma) \le a < S_1+S_3+S(\sigma), S_3+S(\tau) > b,\ \square\}$$

$$= \{S_1+S_3+S(\sigma) > a,\ S_3 + S(\tau) > b,\ \square\} - \{S_1+S(\sigma) > a, S_3+S(\tau) > b,\ \square\}.$$

Substituting this in (*) yields

$$(6.1) \qquad P\{S_1+S_2+S(\sigma) > a,\ S_1+S_3+S(\tau) > b,\ \square\}$$

$$= P\{S_1+S(\sigma) > a,\ S_1+S_3+S(\tau) > b,\ \square\}$$

$$+ P\{S_1+S_3+S(\sigma) > a,\ S_3+S(\tau) > b,\ \square\}$$

$$- P\{S_1+S(\sigma) > a,\ S_3+S(\tau) > b,\ \square\}$$

with $(\sigma \cup \tau) \cap \{1,2,3\} = \phi$ and $\square$ consisting of inequalities not involving $S_1, S_2, S_3$ except through the combination $S_1+S_2+S_3$.

Choose $\xi$ and $\eta$ such that $|\xi-\eta| = i$, $|\eta-\xi| = j$ and $|\xi \cap \eta| = k$. Now define

$$Q(i,j,k) = P\{S(\xi) > a,\ S(\eta) > b\} .$$

Simplify (6.1) by deleting $\square$ to obtain the desired recursion

$$(6.2) \qquad Q(i,j,k) = Q(i-1,j,k) + Q(i,j-1,k) - Q(i,j,k-1) .$$

The boundary terms $Q(0,j,k)$, $Q(i,0,k)$ and $Q(i,j,0)$ are easily evaluated (see below). Thus (6.2) gives an efficient method for tabulating $Q(i,j,k)$. By varying $a$ and $b$ the joint c.d.f. of $S(\xi)$ and $S(\eta)$ may be calculated.

Now the boundary terms are evaluated. Let $\sigma_1, \sigma_2, \ldots, \sigma_m$ and $\{1,2,\ldots,m\}$ be disjoint sets. Define $k_i = |\sigma_i|$ for $1 \le i \le m$ and $k = \Sigma_i k_i$. It is reasonably intuitive (and fairly easy to prove) that

$$(6.3)\qquad P\left(\bigcap_{i=1}^{m}\{S(\sigma_i) \le a_i < S_i + S(\sigma_i)\}\right)$$

$$= \binom{n-1}{k_1,\ldots,k_m}(1-\Sigma_i a_i)_+^{n-1-k} \prod_j a_j^{k_j} .$$

If $\sigma_1 = \cdots = \sigma_m = \phi$, (6.3) reduces to the fundamental lemma (1.1). Let $\underset{\sim}{a} = (a_1,\ldots,a_m)$ and $\underset{\sim}{k} = (k_1,\ldots,k_m)$. Define

$$R(\underset{\sim}{a};\underset{\sim}{k}) = \binom{n-1}{k_1,\ldots,k_m}(1-\Sigma_i a_i)_+^{n-1-k} \prod_j a_j^{k_j}$$

with $k = \Sigma_i k_i$. Every event $\{S(\xi) > b\}$ can be expressed as a disjoint union of events of the form $\{S(\sigma) \le b < S_i + S(\sigma)\}$. For instance

$$\{S_1+S_2+S_3 > b\} = \{S_1 > b\} \cup \{S_1 \le b < S_1+S_2\} \cup \{S_1+S_2 \le b < S_1+S_2+S_3\} .$$

Using this type of expansion and (6.3) one obtains

$$(6.4)\qquad P\left(\bigcap_{i=1}^{m}\{S(\xi_i) > a_i\}\right) = \sum_{\underset{\sim}{\ell} \le \underset{\sim}{k}} R(\underset{\sim}{a};\underset{\sim}{\ell})$$

where $\xi_1,\ldots,\xi_m$ are nonempty and disjoint with $|\xi_i| = 1+k_i$ for $1 \le i \le m$, $\underset{\sim}{k} = (k_1,\ldots,k_m)$, and $\underset{\sim}{\ell} \le \underset{\sim}{k}$ iff $\ell_i \le k_i$ for all $i$. This has the special case

$$Q(p,q,0) = \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} R(a,b;i,j) .$$

Let $\xi_1 \subset \xi_2 \subset \cdots \subset \xi_m$ and $0 \le k_1 < k_2 < \cdots < k_m < n-1$ with $|\xi_i| = k_i+1$ for $1 \le i \le m$. Let $0 = b_0 < b_1 < \cdots < b_m < 1$ and define $a_i = b_i - b_{i-1}$ for $1 \le i \le m$. It is clear by inspection that

$$(6.5) \qquad P\left(\bigcap_{i=1}^{m} \{S(\xi_i) > b_i\}\right) = \sum_{\underset{\sim}{\ell} \in D} R(\underset{\sim}{a};\underset{\sim}{\ell})$$

with $D = \{(\ell_1,\ldots,\ell_m): \sum_{j=1}^{p} \ell_j \le k_p \text{ for } 1 \le p \le m\}$. To see this most easily use exchangeability (1.2) and take $\xi_i = \{1,2,\ldots,k_i+1\}$ for all $i$ without loss of generality. Generate $(S_1,\ldots,S_n)$ as the successive differences between the order statistics of $n-1$ points chosen uniformly and independently on the interval $[0,1]$ (see the proof of (1.1)). Let $L_i$ be the number of points falling in $(b_{i-1},b_i]$. Define $\underset{\sim}{L} = (L_1,\ldots,L_m)$. The event $\bigcap_{i=1}^{m} \{S(\xi_i) > b_i\}$ occurs iff $\underset{\sim}{L} \in D$. (6.5) follows immediately upon using the multinomial distribution of $\underset{\sim}{L}$.

A special case of (6.5) is

$$Q(p,0,q) = \sum_{i=1}^{q-1} \sum_{j=1}^{p+q-1-i} R(b,a-b;i,j)$$

when $b < a$. The degenerate case $b \ge a$ is easily handled. The remaining type of boundary case is $Q(0,p,q)$. This reduces to the previous case since

$$Q(a,b;i,j,k) = Q(b,a;j,i,k) \ .$$

Here the dependence of $Q$ on $a$ and $b$ has been made explicit in the notation in the obvious way.

## 7. Conditional Moments of V.

This section is devoted to developing expressions for some conditional moments of the vacancy $V$. These expressions follow easily from the results of section three.

First some notation will be introduced. This notation is a slight modification of that in section three. Let $I_j$ be the indicator of the event that $P_j$ is not covered. Remember that $P_j$ is the counterclockwise endpoint of the $j^{th}$ arc. Then

$$G = \sum_{j=1}^{n} I_j .$$

Let $\sigma \subset \{1,2,\ldots,n\}$. Then

$$\binom{G}{q} = \sum_{|\sigma|=q} \prod_{j\epsilon\sigma} I_j \tag{7.1}$$

where the sum is over all subsets containing $q$ elements.

Using this formula gives

$$EV^p\binom{G}{q} = \sum_{|\sigma|=q} EV^p \prod_{j\epsilon\sigma} I_j = \binom{n}{q} EV^p I_1 I_2 \cdots I_q .$$

Therefore

$$E(V^p | P_1, P_2, \ldots, P_q \text{ are uncovered}) = \frac{EV^p\binom{G}{q}}{E\binom{G}{q}} . \tag{7.2}$$

The numerator and denominator are given by formulas (3.3) and (3.1) respectively. Remember that the $P_i$ are the unordered endpoints.

An expression for $E(V^p|G=k)$ will now be derived. The argument uses the following combinatorial identity:

$$(7.3) \qquad I_{\{G=k\}} = \sum_{j \geq k} (-1)^{j-k} \binom{j}{k} \binom{G}{j} .$$

This identity will be verified by manipulating the indicator functions $I_j$ defined previously. Elementary inclusion-exclusion arguments may also be given.

$$I_{\{G=k\}} = \sum_{|\sigma|=k} \Big( \prod_{i \in \sigma} I_i \Big) \Big( \prod_{i \notin \sigma} (1-I_i) \Big) .$$

Expanding the product $\prod_{i \notin \sigma} (1-I_i)$ yields

$$= \sum_{|\sigma|=k} \sum_{\pi \supset \sigma} (-1)^{|\pi-\sigma|} \prod_{i \in \pi} I_i .$$

Now group terms to get

$$= \sum_{|\pi| \geq k} (-1)^{|\pi|-k} \binom{|\pi|}{k} \prod_{i \in \pi} I_i$$

$$= \sum_{j \geq k} (-1)^{j-k} \binom{j}{k} \sum_{|\pi|=j} \prod_{i \in \pi} I_i .$$

Now using (7.1) completes the derivation of (7.3).

Taking expectations on both sides of (7.3) yields

$$(7.4) \qquad P\{G=k\} = \sum_{j \geq k} (-1)^{j-k} \binom{j}{k} E\binom{G}{j} .$$

The distribution of $G$ was first obtained by Holst and Siegel. The desired conditional moments are now easily obtained.

$$(7.5) \quad E(V^p|G=k) = \frac{E\,V^p\,I_{\{G=k\}}}{P\{G=k\}} = \frac{1}{P\{G=k\}} \sum_{j \geq k} (-1)^{j-k} \binom{j}{k} E\,V^p \binom{G}{j}$$

upon using (7.3). Summing over $k$ in (7.3) and using an elementary property of binomial coefficients one obtains the related identity

$$I_{\{G \geq k\}} = \sum_{j \geq k} (-1)^{j-k} \binom{j-1}{k-1} \binom{G}{j} .$$

This can be used to obtain expressions for $E(V^p|G \geq k)$ .

Formula (7.5) is a summation of terms involving $E\,V^p\binom{G}{q}$ where $q$ varies. To avoid possible misinterpretation of the spacings $S_j$ formula (3.3) will be restated with the dependence on $p$ and $q$ made more explicit in the notation.

$$E\,V^p\binom{G}{q} = \binom{n}{q} E \prod_{k=1}^{q} F(S_{kr}) \{ \sum_{j=1}^{r} g(S_{jr}) \}^{n-q}$$

where $r = p+q$ and $S_{1r}, S_{2r}, \ldots, S_{rr}$ are the ordered spacings between $r$ points chosen uniformly at random on the circumference. $g(s) \equiv \int_0^s F(x)dx$.

## 8. An Upper Bound for the Vacancy.

As in section (0) let $P_1, P_2, \ldots, P_n$ be the unordered endpoints of arcs of length $L_1, L_2, \ldots, L_n$ which are i.i.d. from $F$. Let $S_i$ be the distance from $P_i$ to the nearest point $P_j$ in the clockwise direction. This is a convenient modification of the notation in section (0). Due to the exchangeability of the spacings the joint distribution of $S_1, S_2, \ldots, S_n$ still satisfies the fundamental lemma (1.1).

Define the random variable $V^*$ as

$$V^* = \sum_{i=1}^{n} (S_i - L_i)_+ . \tag{8.1}$$

Comparing this with (2.4) shows that $V = V^*$ when the arcs are of fixed length $a$ ($F(x) = I_{\{x \geq a\}}$). In general $V \leq V^*$. This is easily seen as follows. Let $x$ be a point on the circumference. Let $P_j$ be the first endpoint which is reached upon starting from $x$ and moving counterclockwise. The $j^{\underline{th}}$ arc will be said to be the arc immediately preceding $x$. $V^*$ can now be described as the measure of the set of points which are not covered by the arc immediately preceding them. If a point $x$ is not covered by any of the arcs then in particular $x$ is not covered by the immediately preceding arc. Therefore $V \leq V^*$ as desired.

Expressions for the distribution and moments of $V^*$ are readily obtained. The argument of Holst will be used to find the distribution of $V^*$ (see the second derivation of the moments of $V$ in section two). Let $H$ be an arbitrary function. Define $I_j = I_{\{S_j > L_j\}}$ for $1 \leq j \leq n$. Using exchangeability it follows that

$$EH(V^*) = \sum_{k=0}^{n} \binom{n}{k} EH\left(\sum_{i=1}^{k} (S_i - L_i)\right) \prod_{i=1}^{k} I_i \prod_{i=k+1}^{n} (1-I_i)$$

$$= \sum_{k=0}^{n} \sum_{j=0}^{n-k} \binom{n}{k}\binom{n-k}{j}(-1)^j \; EH\left(\sum_{i=1}^{k} (S_i - L_i)\right) \prod_{i=1}^{k+j} I_i \; .$$

To evaluate $EH\left(\sum_{i=1}^{k} (S_i - L_i)\right) \prod_{i=1}^{k+j} I_i$ condition on the values of $L_1, L_2, \ldots, L_n$ and use (1.1) and (1.4). This yields

$$= E\left(1 - \sum_{i=1}^{k+j} L_i\right)_+^{n-1} E\left\{H\left(\left(1 - \sum_{i=1}^{k+j} L_i\right)\left(\sum_{\ell=1}^{k} S_\ell\right)\right) \Big| L_1, \ldots, L_n\right\}$$

$$(\dagger) \qquad = \int_0^1 (1-y)^{n-1} dF^{(k+j)*}(y) \frac{\Gamma(n)}{\Gamma(k)\Gamma(n-k)} \int_0^1 H((1-y)s) s^{k-1} (1-s)^{n-k-1} ds$$

since $\sum_{\ell=1}^{k} S_\ell \sim \text{Beta}(k,n-k)$. $F^{(k)*}$ is defined to be the k-fold convolution: $F^{(k)*}(x) = P\{\sum_{i=1}^{k} L_i \le x\}$. Making the transformation $u = (1-y)s$ and interchanging the order of integration this becomes

$$\frac{\Gamma(n)}{\Gamma(k)\Gamma(n-k)} \int_0^1 H(u) u^{k-1} du \int_0^{1-u} (1-u-y)^{n-k-1} dF^{(k+j)*}(y) \; .$$

Note that when $k = 0$ or $k = n$ the Beta distribution is degenerate. After separating out these special cases the final result becomes

$$(8.2) \qquad EH(V^*) = \int_0^1 H(1-y)(1-y)^{n-1} dF^{(n)*}(y)$$

$$+ H(0) \sum_{j=0}^{n} \binom{n}{j} (-1)^j \int_0^1 (1-y)^{n-1} dF^{(j)*}(y)$$

$$+ \sum_{k=1}^{n-1} \sum_{j=0}^{n-k} \binom{n}{k}^2 \frac{k(n-k)}{n} \binom{n-k}{j} (-1)^j \int_0^1 H(u) u^{k-1} du \int_0^{1-u} (1-u-y)^{n-k-1} dF^{(k+j)*}(y) \; .$$

If $F^{(n)*}$ is absolutely continuous with density $f_n$, then (apart from an atom at zero) the distribution of $V^*$ is absolutely continuous with density $q(u)$ given by

$$(8.3) \qquad q(u) = w\delta_0 + u^{n-1} f_n(1-u)$$

$$+ \sum_{k=1}^{n-1} \sum_{j=0}^{n-k} \binom{n}{k}^2 \frac{k(n-k)}{n} \binom{n-k}{j} (-1)^j \, u^{k-1} \int_0^{1-u} (1-u-y)^{n-k-1} dF^{(k+j)*}(y) \; .$$

$w$ is the mass of the atom at zero,

$$w = \sum_{j=0}^{n} \binom{n}{j} (-1)^j \int_0^1 (1-y)^{n-1} dF^{(j)*}(y) \; .$$

(8.3) is obtained by inspection from (8.2).

To obtain the cumulative distribution of $V^*$ set $H(u) = I_{\{u \le x\}}$ in (8.2). Using the fact that

$$P\{\sum_{\ell=1}^{k} S_\ell \le x\} = \sum_{j=k}^{n-1} \binom{n-1}{j} x^j (1-x)^{n-1-j}$$

for $0 \le x \le 1$ and the earlier expression (†) the following result may be derived:

$$(8.4) \qquad P\{V^* \le x\} = 1 + \sum_{i=1}^{n} \sum_{k=0}^{i-1} \binom{n}{i} \binom{n-1}{k} \binom{i-1}{k} (-1)^{i+k} \, x^k \int_0^{1-x} (1-x-y)^{n-k-1} dF^{(i)*}(y).$$

The moments of $V^*$ may be found by taking $H(x) = x^p$ in (8.2) or by duplicating the first or third derivation in section two and conditioning on $L_1, L_2, \ldots, L_n$ at the appropriate step. This leads to

$$(8.5) \qquad E(V^*)^p = \binom{p+n-1}{p}^{-1} \sum_{k=1}^{p} \binom{n}{k}\binom{p-1}{k-1} \int_0^1 (1-y)^{n+p-1} dF^{(k)*}(y) \ .$$

Observe that (8.4) and (8.5) reduce immediately to the results of Siegel upon substituting $F(x) = I_{\{x \geq a\}}$.

$V \leq V^*$ so that $P\{V \leq x\} \geq P\{V^* \leq x\}$. Some idea of the error in approximating $P\{V \leq x\}$ by $P\{V^* \leq x\}$ may be gained by the comparison of $EV^p$ and $E(V^*)^p$ which are known quantities. The approximation will not be good unless $F$ is tightly concentrated about its mean. For example let $F$ be the uniform distribution on the interval $[\alpha-\beta, \beta+\alpha]$ with $0 < \beta < \alpha < 1-\beta$. Using (3.2) and (8.5) gives

$$EV = (1-EL)^n = (1-\alpha)^n \quad \text{and}$$

$$EV^* = E(1-L)^n = \frac{(1-\alpha+\beta)^{n+1} - (1-\alpha-\beta)^{n+1}}{2\beta(n+1)} \ .$$

Since $V \leq V^*$ it is reasonable to use

$$\frac{EV^* - EV}{EV} = \frac{(1-\alpha+\beta)^{n+1} - (1-\alpha-\beta)^{n+1}}{2\beta(n+1)(1-\alpha)^n} - 1$$

as a measure of the similarity of the distributions of $V$ and $V^*$.

This quantity is large when $\beta$ is large. $(EV^*-EV)/EV = \frac{2^n}{n+1} - 1$ when $\alpha = \beta = \frac{1}{2}$. However $\frac{EV^*-EV}{EV} \sim \frac{n(n-1)}{6} (\frac{\beta}{1-\alpha})^2$ as $\beta \downarrow 0$. Thus the approximation is quite good for small $\beta$.

The argument leading to (8.2) does not use the fact that $L_1, L_2, \ldots, L_n$ are i.i.d. but only that they are exchangeable. Thus (8.2) through (8.5) remain true under this weaker hypothesis. It is necessary only to assume that $L_1, L_2, \ldots, L_n$ are exchangeable and independent of $S_1, S_2, \ldots, S_n$. Note that $F^{(k)*}(x) \equiv P\{\Sigma_{i=1}^k L_i \leq x\}$ can no longer be interpreted as a convolution.

## 9. The Largest Gap.

The vacancy consists of $G$ disjoint gaps. Let $H$ denote the length of the largest of these gaps. This section will be concerned primarily with the distribution of $H$.

Increase the length of each arc by $t$ while keeping the counterclockwise endpoint fixed. The $j$-th arc now has length $L_j + t$. Let $G(t)$ and $H(t)$ be respectively the number of uncovered gaps and the length of the largest gap in this new configuration. By definition $G(0) = G$ and $H(0) = H$.

Clearly $H(t) = (H-t)_+$. Thus $H \leq t$ if and only if $H(t) = 0$ or equivalently $G(t) = 0$. This yields

$$(9.1) \qquad P\{H \leq t\} = P\{G(t) = 0\} .$$

The probability of complete coverage is given by (7.4) as

$$P\{G(t)=0\} = 1 + \sum_{j=1}^{n} (-1)^j E\binom{G(t)}{j} \ . \tag{9.2}$$

Define $F_t$ to be the cumulative distribution function of $L_i + t$ so that $F_t(x) = F(x-t)$. Correspondingly define

$$g_t(s) \equiv \int_0^s F_t(x)dx = \int_0^{s-t} F(x)dx = g(s-t) \ .$$

Note that if $s < t$ then $g_t(s) = g(s-t) = 0$. Now replace $F$ and $g$ in (3.1) by $F_t$ and $g_t$ to obtain

$$E\binom{G(t)}{q} = \binom{n}{q} E\left(\prod_{i=1}^{q} F(S_i - t)\right)\left\{\sum_{k=1}^{q} g(S_k - t)\right\}^{n-q} \tag{9.3}$$

with all terms defined as in (3.1). The distribution $F_t$ may have $F_t(1) < 1$ but this does not affect the validity of (9.3) because the general results (3.1), (3.2) and (3.3) continue to hold when $F(1) < 1$.

Combining (9.1), (9.2) and (9.3) immediately gives an expression for the distribution of $H$.

Define $H_k$ to be the length of the k-th largest uncovered gap so that $H = H_1 \geq H_2 \geq \cdots \geq H_G$ and $\Sigma_{k=1}^{G} H_k = V$. For convenience set $H_k = 0$ for $k > G$. The distribution of $H_k$ is found in the same way as that of $H_1$. For $t \geq 0$,

$$P\{H_k > t\} = P\{G(t) \geq k\} = \sum_{j \geq k} (-1)^{j-k} \binom{j-1}{k-1} E\binom{G(t)}{j}$$

by the formula following (7.5). An application of (9.3) completes the result.

In the remainder of this section a simple upper bound for the moments of $H$ will be developed. This will yield an upper bound for $P\{H > t\}$.

Choose points $Q_1, Q_2, \ldots, Q_p$ distributed uniformly and independently on the circumference and independent of the $n$ random arcs. Let $B$ be the event that all the points $Q_i$ lie in the same gap. One way to calculate $P(B)$ is to condition on $P_1, P_2, \ldots, P_n$ and $L_1, L_2, \ldots, L_n$ so that the arcs may be regarded as fixed. The conditional probability that all $p$ of the points $Q_i$ land in the $j$-th largest gap is $H_j^p$. Therefore

$$P(B \mid P_1, \ldots, P_n, L_1, \ldots, L_n) = \sum_j H_j^p .$$

Taking expectations yields

$$P(B) = E \sum_j H_j^p .$$

Another way to calculate $P(B)$ is to condition on the location of the points $Q_1, \ldots, Q_p$. Let $S_{1p}, S_{2p}, \ldots, S_{pp}$ be the spacings between the points $Q_1, \ldots, Q_p$. The event $B$ occurs if and only if all $n$ arcs lie entirely between the same two adjacent points in $\{Q_i\}_{i=1}^p$. The probability that all $n$ arcs lie entirely in an interval of length $s$ is $g(s)^n$ by the arguments in section three. Therefore $P(B \mid Q_1, \ldots, Q_p) = \Sigma_{i=1}^p g(S_{ip})^n$. Taking expectations and using the exchangeability of $S_{1p}, S_{2p}, \ldots, S_{pp}$ gives $P(B) = pE\, g(S_{1p})^n$.

Equating the expressions for $P(B)$ leads to

$$E\Big(\sum_j H_j^p\Big) = pE\, g(S_{1p})^n . \tag{9.4}$$

If the length of each arc is increased by $t$ while keeping the counterclockwise endpoint fixed, the length of the j-th largest gap becomes $(H_j-t)_+$. The distribution of arc lengths is now $F_t$ and $g$ is replaced by $g_t$. Thus (9.4) is transformed into

$$E \sum_j (H_j-t)_+^p = pE\, g(S_{1p}-t)^n \,. \tag{9.5}$$

An upper bound for the moments of $H = H_1$ follows immediately:

$$E(H-t)_+^p \leq pE\, g(S_{1p}-t)^n \tag{9.6}$$

where

$$E\, g(S_{1p}-t)^n = \begin{cases} g(1-t)^n & \text{for } p = 1\,, \\ (p-1) \int_t^1 g(u-t)^n (1-u)^{p-2}\, du & \text{for } p \geq 2 \end{cases} \tag{9.7}$$

since $S_{1p} \sim \text{Beta}(1,p-1)$.

If $p$ is sufficiently large the summation $\Sigma_j (H_j-t)_+^p$ will tend to be dominated by its largest term $(H_1-t)_+^p$. If $t$ is sufficiently large most of the terms in $\Sigma_j (H_j-t)_+^p$ will be zero and the term $(H_1-t)_+^p$ will again be dominant. Thus when $p$ or $t$ is sufficiently large $E(H-t)_+^p \approx E\, \Sigma_j (H_j-t)_+^p$ and the inequality (9.6) will be fairly tight.

An upper bound for $P\{H \geq x\}$ is obtained by Chebyshev's inequality:

$$P\{H \geq x\} = P\{H-t \geq x-t\} \leq \frac{E(H-t)_+^p}{(x-t)^p}$$

for $0 \leq t < x$. Now using (9.6) gives

$$(9.8) \qquad P\{H \geq x\} \leq \inf_{p,t} \frac{pE\, g(S_{1p}-t)^n}{(x-t)^p}$$

where the infimum is over all integers $p \geq 1$ and all $t$ in $[0,x)$. This upper bound may be approximated numerically using (9.7).

Equation (9.5) mav be given another interpretation. Let $Z_i$ be the length of the uncovered gap having $P_i$ as an endpoint. If $P_i$ is covered by some arc set $Z_i = 0$. Clearly $Z_1, Z_2, \ldots, Z_n$ are exchangeable with $\Sigma_{i=1}^n Z_i = V$. $Z_1, \ldots, Z_n$ is just a reordering of the numbers $H_1, \ldots, H_n$ so that $\Sigma_j (Z_j - t)_+^p = \Sigma_j (H_j - t)_+^p$. From (9.5) and exchangeability it follows that

$$(*) \qquad E(Z_1 - t)_+^p = \frac{p}{n} E\, g(S_{1p} - t)^n \ .$$

This result may be obtained directly from the distribution of $Z_1$. Let $C = \{L_1 < 1\}$ and $D$ be the event that none of the $n-1$ arcs labelled $2,3,\ldots,n$ cover the point $P_1$. $P\{Z_1 > 0\} = P\{P_1 \text{ is uncovered}\} = P(C \cap D) = P(C)P(D) = F(1)g(1)^{n-1}$. Lengthening the arcs by $t$ leads to the similar result

$$(9.9) \qquad P\{Z_1 > t\} = F_t(1) g_t(1)^{n-1} = F(1-t) g(1-t)^{n-1} \ .$$

For $t = 0$ $(*)$ is obtained as follows.

$$EZ_1^p = \int_0^1 u^p P\{Z_1 \varepsilon\, du\} = p \int_0^1 u^{p-1} P\{Z_1 > u\} du$$

$$= p \int_0^1 u^{p-1} F(1-u) g(1-u)^{n-1} du$$

$$= \begin{cases} \dfrac{g(1)^n}{n} & \text{for } p = 1 , \\[2ex] \dfrac{p(p-1)}{n} \displaystyle\int_0^1 u^{p-2} g(1-u)^n du & \text{for } p \geq 2 \end{cases}$$

since $g'(x) = F(x)$. Comparing this with (9.7) completes the result. (*) for $t > 0$ may be verified in the same way.

## 10. The Distribution of V.

In this section a procedure for calculating the distribution of $V$ is outlined. For the distributions $F(x) \sim (x-\alpha)_+^\beta$ considered in section 3 the distribution of $V$ will be found explicitly.

As in section 9 define $Z_i$ to be the length of the uncovered gap having $P_i$ as its clockwise endpoint. If $P_i$ is covered by some arc set $Z_i = 0$. Then $Z_1, Z_2, \ldots, Z_n$ are exchangeable and $\Sigma_i Z_i = V$. The joint distribution of $Z_1, Z_2, \ldots, Z_n$ will now be obtained. In particular it will be shown that

$$(10.1) \qquad P\{Z_1 > t_1, Z_2 > t_2, \ldots, Z_k > t_k\}$$

$$= P\{Z_1 > \sum_{i=1}^k t_i,\ Z_2 > 0, \ldots, Z_k > 0\}$$

$$= (1-x)^{k-1} E \prod_{i=1}^k F((1-x)S_{ik})\{\sum_{j=1}^k g((1-x)S_{jk})\}^{n-k}$$

where $x = \Sigma_{i=1}^k t_i$ and as usual $S_{1k}, S_{2k}, \ldots, S_{kk}$ are the spacings between $k$ independent points uniformly distributed on the circumference and $g(s) = \int_0^s F(x)dx$. Equation (10.1) is a generalization of the fundamental lemma (1.1). Equation (10.1) reduces to (1.1) upon taking $F(x) \equiv 1$ and $g(x) = x$. This means the arcs have been shrunk down to points.

The proof of (10.1) is very similar to that of (3.1). First some notation must be developed. For $1 \le i \le k$ let $Q_i$ be the point obtained by starting from $P_i$ and moving a distance $t_i$ in the counterclockwise direction. Then in the notation of section 4 $[Q_i, P_i]$ is a segment of the circumference with counterclockwise endpoint $Q_i$, clockwise endpoint

$P_i$ and length $t_i$. Let $S_{1k}, S_{2k}, \ldots, S_{kk}$ be the spacings between the points $Q_1, Q_2, \ldots, Q_k$. More precisely, for $1 \leq i \leq k$ define $S_{ik}$ to be the distance from $Q_i$ to the nearest point $Q_j$ in the clockwise direction. $Q_1, Q_2, \ldots, Q_k$ are independent and uniformly distributed so that $S_{1k}, S_{2k}, \ldots, S_{kk}$ have the usual joint distribution for the spacings.

Define $B = \{Z_1 > t_1, Z_2 > t_2, \ldots, Z_k > t_k\}$. Let $A_i$ denote the arc with counterclockwise endpoint $P_i$ and length $L_i$. The event $B$ occurs if and only if $A_1, \ldots, A_n$ do not intersect the segments $[Q_1, P_1], \ldots, [Q_k, P_k]$. More precisely, $B$ occurs if and only if $\bigcup_{i=1}^{n} A_i$ and $\bigcup_{i=1}^{k} [Q_i, P_i]$ are disjoint. Let $C$ be the event that $\bigcup_{i=1}^{k} A_i$ and $\bigcup_{i=1}^{k} [Q_i, P_i]$ are disjoint and $D$ be the event that $\bigcup_{i=k+1}^{n} A_i$ and $\bigcup_{i=1}^{k} [Q_i, P_i]$ are disjoint. Then $B = C \cap D$.

Now condition on the points $P_1, P_2, \ldots, P_k$ so that $[Q_1, P_1], \ldots, [Q_k, P_k]$ can be regarded as fixed in the argument which follows. $P(B|*) = P(C|*)P(D|*)$ where $*$ denotes conditioning on $P_1, P_2, \ldots, P_k$. This follows from the mutual independence of $L_1, \ldots, L_n$, $P_1, \ldots, P_n$. $C$ occurs if and only if $L_i \leq S_{ik} - t_i$ for $1 \leq i \leq k$. Thus

$$P(C|*) = \prod_{i=1}^{k} F(S_{ik} - t_i) \ .$$

Since the arcs $A_{k+1}, \ldots, A_n$ are conditionally independent $P(D|*) = P(E|*)^{n-k}$ where $E$ is the event that $A_n$ and $\bigcup_{i=1}^{k} [Q_i, P_i]$ are disjoint. Following the arguments of section 3 will lead to

$$P(E|*) = \sum_{i=1}^{k} g(S_{ik} - t_i) \ .$$

Putting this all together gives

$$P(B|*) = \prod_{i=1}^{k} F(S_{ik}-t_i)\{\sum_{j=1}^{k} g(S_{jk}-t_j)\}^{n-k} .$$

Taking expectations yields

$$P\{Z_1 > t_1,\ldots,Z_k > t_k\}$$

$$= E \prod_{i=1}^{k} F(S_{ik}-t_i)\{\sum_{j=1}^{k} g(S_{jk}-t_j)\}^{n-k}$$

$$= P(W)E(\prod_{i=1}^{k} F(S_{ik}-t_i)\{\sum_{j=1}^{k} g(S_{jk}-t_j)\}^{n-k}|W)$$

where $W = \bigcap_{i=1}^{k} \{S_{ik} > t_i\}$. Applying (1.1) and (1.4) to this expression completes the proof.

When $F(x) = x^{\beta}$ so that $g(x) = \frac{x^{\beta+1}}{\beta+1}$ equation (10.1) becomes

(10.2) $$P\{Z_1 > t_1, Z_2 > t_2,\ldots,Z_k > t_k\} = \xi_k(\beta)(1 - \sum_{i=1}^{k} t_i)_+^{n(\beta+1)-1}$$

where

$$\xi_k(\beta) = E(\prod_{i=1}^{k} S_{ik}^{\beta})\left\{\sum_{j=1}^{k} \frac{S_{jk}^{\beta+1}}{\beta+1}\right\}^{n-k} = P\{Z_1 > 0, Z_2 > 0,\ldots,Z_k > 0\} .$$

The quantities $\xi_k(\beta)$ were evaluated by Holst and Siegel (1982) and can also be obtained as a special case of (3.4).

An expression for the distribution of $V$ may be found by the argument of Holst (1980) used in sections two and eight. For $1 \leq k \leq n$ define $Y_k = \Sigma_{k=1}^{k} Z_i$, $I_k = I_{\{Z_k > 0\}}$ and $B_k = \bigcap_{i=1}^{k} \{Z_i > 0\}$. Note that $Y_n = V$. Then by exchangeability of $Z_1, Z_2, \ldots, Z_n$:

$$
\begin{aligned}
P\{V > x\} &= \sum_{j=1}^{n} \binom{n}{j} EI_{\{Y_j > x\}} \prod_{i=1}^{j} I_i \prod_{i=j+1}^{n} (1-I_i) \\
&= \sum_{j=1}^{n} \binom{n}{j} \sum_{k=j}^{n} (-1)^{k-j} \binom{n-j}{k-j} EI_{\{Y_j > x\}} \prod_{i=1}^{k} I_i
\end{aligned}
$$

so that

$$
(10.3) \qquad P\{V > x\} = \sum_{j=1}^{n} \binom{n}{j} \sum_{k=j}^{n} (-1)^{k-j} \binom{n-j}{k-j} P\{Y_j > x | B_k\} P(B_k) \quad .
$$

From (10.1)

$$
P(B_k) = E\left(\prod_{i=1}^{k} F(S_{ik})\right)\left\{\sum_{j=1}^{k} g(S_{jk})\right\}^{n-k} = \binom{n}{k}^{-1} E\binom{G}{k}
$$

using (3.1). Let $F_k(t_1, t_2, \ldots, t_k) dt_1 dt_2 \cdots dt_k = P\{X_1 \varepsilon dt_1, \ldots, Z_k \varepsilon dt_k | B_k\}$ so that $f_k$ is the joint density of $Z_1, Z_2, \ldots, Z_k$ given $B_k$. $f_k$ may by calculated from (10.1) by partial differentiation. One could obtain $P\{Y_j > x | B_k\}$ for $j \leq k$ by integrating $f_k(t_1, \ldots, t_k)$ over the set where $t_1 + t_2 + \cdots + t_j > x$. Thus in principle (10.1) and (10.3) together determine the distribution of $V$.

The distribution of $V$ may be given explicitly when $F(x) = x^{\beta}$. In this case (10.2) yields

$$
P\{Z_1 > t_1, \ldots, Z_k > t_k | B_k\} = \left(1 - \sum_{i=1}^{k} t_i\right)_+^{n(\beta+1)-1}
$$

and thus

$$f_k(t_1,t_2,\ldots,t_k) \propto \left(1 - \sum_{i=1}^{k} t_i\right)_+^{n(\beta+1)-k-1}$$

which is the density of a Dirichlet distribution. This implies

$$\mathcal{L}(Y_j | B_k) = \text{Beta}(j, n(\beta+1)-j) \quad \text{for} \quad j \leq k .$$

Note the curious fact that $\mathcal{L}(Y_j | B_k)$ does not depend on $k$ so long as $j \leq k$. Define $H_j(x) = P\{Y_j > x | B_k\}$ when $j \leq k$. $P(B_k) = \xi_k(\beta)$. Thus (10.3) becomes

$$\text{(10.4)} \qquad P\{V > x\} = \sum_{j=1}^{n} H_j(x) \binom{n}{j} \sum_{k=j}^{n} (-1)^{k-j} \binom{n-j}{k-j} \xi_k(\beta)$$

$$= \sum_{j=1}^{n} H_j(x) P\{G=j\}$$

by using (7.4). This may be restated as

$$\text{(10.5)} \qquad \mathcal{L}(V | G=k) = \text{Beta}(k, n(\beta+1)-k)$$

when $F(x) = x^\beta$.

The distribution of $V$ can also be obtained when $F(x) = \lambda(x-a)_+^\beta$ where $\lambda$ is such that $F(1) \leq 1$. Only a sketch will be given. It is necessary to evaluate the terms $P(B_k)$ and $P\{Y_j > x | B_k\}$ found in (10.3).

It is easy to show that

$$(\#) \qquad P_\lambda(B_k) = \lambda^n P_1(B_k) \quad \text{and} \quad P_\lambda\{Y_j > x \mid B_k\} = P_1\{Y_j > x \mid B_k\}$$

where the subscript denotes the value of $\lambda$. Therefore it suffices to consider only $\lambda = 1$ in the following.

As usual let $Z_1, Z_2, \ldots, Z_n$ be the gap lengths generated by $n$ random arcs with lengths i.i.d. from the distribution $F(x) = (x-a)_+^\beta$. Let $Z_1^*, Z_2^*, \ldots, Z_n^*$ be the gap lengths generated by $n$ random arcs with lengths i.i.d. from the distribution $F^*(x) = x^\beta$. The argument in section 9 which involves increasing the length of the arcs shows that

$$(\dagger) \qquad \mathcal{L}(Z_1, Z_2, \ldots, Z_n) = \mathcal{L}((Z_1^*-a)_+, (Z_2^*-a)_+, \ldots, (Z_n^*-a)_+) \ .$$

From (10.2) one obtains the following analogue of (1.4) for $1 \le k \le n$:

$$(@) \qquad \mathcal{L}(Z_1^*-t_1, \ldots, Z_k^*-t_k \mid Z_1^* > t_1, \ldots, Z_k^* > t_k)$$
$$= \mathcal{L}((1 - \sum_{i=1}^k t_i)(Z_1^*, \ldots, Z_k^*) \mid Z_1^* > 0, \ldots, Z_k^* > 0) \ .$$

Using ($\dagger$) gives

$$P_1(B_k) = P(\bigcap_{i=1}^k \{Z_i > 0\}) = P(\bigcap_{i=1}^k \{Z_i^* > a\}) = \xi_k(\beta)(1-ka)_+^{n(\beta+1)-1}$$

by (10.2). Define $Y_j = \Sigma_{i=1}^j Z_i$ and $Y_j^* = \Sigma_{i=1}^j Z_i^*$. Then for $j \le k$ equation ($\dagger$) yields

$$P_1\{Y_j > x | B_k\} = P\{Y_j^* > x+ja | Z_1^* > a, \ldots, Z_k^* > a\} .$$

Upon applying (@) this becomes

$$= P\{Y_j^* > \frac{x}{1-ka} | Z_1^* > 0, \ldots, Z_k^* > 0\} .$$

The result preceding (10.4) gives the conditional distribution $Y_j^*$ as Beta$(j,n(\beta+1)-j)$ and thus the above expression equals $H_j(\frac{x}{1-ka})$ where $H(\cdot)$ is the same as the function in (10.4):

$$H_j(y) = \frac{\Gamma(n(\beta+1))}{\Gamma(j)\Gamma(n(\beta+1)-j)} \int_y^1 u^{j-1}(1-u)^{n(\beta+1)-j-1}du .$$

Applying (#) to the previous results produces

$$P(B_k) = \lambda^n \xi_k(\beta)(1-ka)_+^{n(\beta+1)-1} \quad \text{and} \tag{10.6}$$

$$P\{Y_j > x | B_k\} = H_j(\frac{x}{1-ka}) .$$

Plugging these into (10.3) yields the distribution of $V$ when $F(x) = \lambda(x-a)_+^{\beta}$.

## References


Cressie, N. (1977). On some properties of the scan statistic on the circle and the line. Journal of Applied Probability, 14, 272-283.

Fisher, R. A. (1940). On the similarity of the distributions found for the test of significance in harmonic analysis, and in Steven's problem in geometrical probability. Ann. Eugenics, 10, 14-17.

Holst, L. (1980). A note on random arcs on the circle. Technical Report, Department of Statistics, Stanford University.

Holst, L. and Siegel, A. (1982). Covering the circle with random arcs of random sizes. Journal of Applied Probability, to appear.

Jewell, N. and Romano, J. (1982). Coverage problems and random convex hulls. Journal of Applied Probability, to appear.

Naus, J. (1966). A power comparison of two tests of non-random clustering. Technometrics, 8, 493-517.

Robbins, H. E. (1944). On the measure of a random set. Ann. Math. Stat., 15, 70-74.

Siegel, A. (1977). Random coverage problems in geometrical probability with an application to time series analysis. Ph.D. dissertation, Department of Statistics, Stanford University.

Siegel, A. (1978a). Random space filling and moments of coverage in geometrical probability. Journal of Applied Probability, 15, 340-355.

Siegel, A. (1978b). Random arcs on the circle. Journal of Probability, 15, 774-789.

Siegel, A. (1979). The noncentral chi-squared distribution with zero degrees of freedom and testing for uniformity. Biometrika, 66, 381-386.

Solomon, H. (1978). Geometric Probability. SIAM: Philadelphia.

Stevens, W. L. (1939). Solution to a geometrical problem in probability. Ann. Eugenics, 9, 315-320.

Votaw, D. F., Jr. (1946). The probability distribution of the measure of a random linear set. Ann. Math. Stat., 17, 240-244.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER 326 | 2. GOVT ACCESSION NO. AD-A120660 | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) The Moments and Distributions of some Quantities Arising from Random Arcs on the Circle | | 5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s) FRED HUFFER | | 8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0475 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS Office Of Naval Research Statistics & Probability Program Code 411SP Arlington, VA 22217 | | 12. REPORT DATE September 28, 1982 |
| | | 13. NUMBER OF PAGES 79 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 15. SECURITY CLASS. (of this report) UNCLASSIFIED |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Random Arcs on the Circle,
Distribution of uncovered gaps and circle covered,
Moments and joint moments

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE REVERSE SIDE.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

#326

THE MOMENTS AND DISTRIBUTIONS OF SOME QUANTITIES ARISING FROM RANDOM ARCS ON THE CIRCLE

Fred W. Huffer,

Consider the random uniform placement of $n$ arcs on the circle where the arc lengths are sampled from a distribution $F$ on $[0,1]$. Let $G$ be the number of uncovered gaps on the circumference and $V$ be the amount of the circle which is not covered. General expressions are given for the moments, joint moments and distributions of $G$ and $V$. These expressions are evaluated for distributions of the form $F(x) = (x-a)_+^\beta$. Let $G_k$ denote the number of segments on the circle which are covered exactly $k$ times and $V_k$ denote the measure of the region covered exactly $k$ times. Formulas are derived for the first and second order moments of $G_k$ and $V_k$ for all $k$. Also, the distribution of $H_k$ is obtained where $H_k$ is the length of the $k^{th}$ largest uncovered gap on the circumference. Additional derivations of many of these results are provided for the cases $F(x) = x$ and $F(x) = I_{\{x \geq a\}}$ which possess special properties.

S N 0102- LF- 014- 6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED